GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 04:20:59 ; Search time 1474.31 Seconds
(without alignments)
138.851 Million cell updates/sec

Title: PCT-US00-32259-12
Perfect score: 40
Sequence: 1 ctcgagccaccccataaccctcaatactccagggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vi:*
29: gb_htg1:*
30: gb_htg2:*
31: gb_in1:*
32: gb_in2:*
33: em_ba1:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: em_htg1:*
44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_htg8:*
50: gb_htg9:*
51: gb_htg10:*
52: gb_htg11:*
53: gb_htg12:*
54: gb_htg13:*
55: gb_htg14:*
56: gb_in3:*
57: gb_htg15:*
58: gb_htg16:*
59: gb_htg17:*
60: em_htg4:*
61: em_htg5:*
62: em_htg6:*
63: em_htg7:*
64: em_hum6:*
65: gb_htg18:*
66: gb_htg19:*
67: gb_htg20:*
68: gb_htg21:*
69: gb_htg22:*
70: gb_htg23:*
71: gb_vi1:*
72: gb_vi2:*
73: gb_ba3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	23	57.5	1653	37	AK026605	AK026605 Homo sapi
c	2	23	57.5	2600	78	HSTEGT	X75861 H.sapiens T
c	3	23	57.5	2634	81	AR058920	AR058920 Sequence
	4	23	57.5	2634	81	AR058921	AR058921 Sequence
	5	23	57.5	197189	42	AC019168	AC019168 Homo sapi
	6	22.8	57.0	171185	70	AP002424	AP002424 Homo sapi
	7	22.8	57.0	177097	70	AP001569	AP001569 Homo sapi
c	8	22.8	57.0	179726	9	AC007052	AC007052 Homo sapi
	9	22.8	57.0	188357	37	AL159996	AL159996 Human DNA
	10	22.8	57.0	200774	70	AP001592	AP001592 Homo sapi
	11	22.4	56.0	151514	77	HS471M13	Z97198 Human DNA s
	12	22	55.0	452	79	G22123	G22123 human STS W
c	13	22	55.0	34887	56	TBR012199	AJ012199 Trypanoso
c	14	22	55.0	76727	77	HS821D21	AL021453 Human DNA
c	15	22	55.0	110387	84	HSAJ9621	AJ009611 Homo sapi
c	16	22	55.0	177540	9	AC006538	AC006538 Homo sapi
c	17	21.8	54.5	154599	41	AC016997	AC016997 Homo sapi
c	18	21.8	54.5	155907	49	AC020991	AC020991 Homo sapi
	19	21.6	54.0	37470	48	CNS00YVD	AL096810 Homo sapi
c	20	21.6	54.0	104436	77	HS611N7	AL035663 Human DNA
	21	21.6	54.0	168925	68	AL356502	AL356502 Homo sapi

c	22	21.6	54.0	208454	69	AL445199	AL445199 Homo sapi
	23	21.6	54.0	225653	69	AL359920	AL359920 Homo sapi
	24	21.2	53.0	70594	40	AC015786	AC015786 Homo sapi
c	25	21.2	53.0	125219	69	AL445228	AL445228 Homo sapi
	26	21.2	53.0	147009	29	AC007618	AC007618 Homo sapi
c	27	21.2	53.0	181215	49	AC021300	AC021300 Homo sapi
	28	21.2	53.0	208992	55	AC061709	AC061709 Homo sapi
c	29	21.2	53.0	233454	9	AC005703	AC005703 Homo sapi
	30	21	52.5	2201	11	AF109155	AF109155 Mus muscu
c	31	21	52.5	16332	1	AF104912	AF104912 Escherich
	32	21	52.5	80514	8	AC004454	AC004454 Homo sapi
c	33	21	52.5	87235	50	AC022963	AC022963 Homo sapi
	34	21	52.5	92215	41	AC018277	AC018277 Drosophil
	35	21	52.5	111269	30	AC009251	AC009251 Drosophil
	36	21	52.5	127409	37	AL137790	AL137790 Human DNA
	37	21	52.5	132571	69	AL445245	AL445245 Homo sapi
	38	21	52.5	155278	78	HSDJ655C4	AL080286 Human DNA
c	39	21	52.5	167277	52	AC025318	AC025318 Homo sapi
	40	21	52.5	167356	70	AP002814	AP002814 Homo sapi
	41	21	52.5	169480	65	AC079608	AC079608 Homo sapi
	42	21	52.5	173333	68	AL355990	AL355990 Homo sapi
	43	21	52.5	181233	38	AC010535	AC010535 Homo sapi
	44	21	52.5	186716	41	AC018468	AC018468 Homo sapi
c	45	21	52.5	200627	42	AC019225	AC019225 Homo sapi

ALIGNMENTS

```
RESULT   1
AK026605/c
LOCUS       AK026605     1653 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ22952 fis, clone KAT09742.
ACCESSION   AK026605
VERSION     AK026605.1  GI:10439494
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens signet-ring cell carcinoma cell_line:KATO III cDNA to
            mRNA, clone_lib:KAT clone:KAT09742.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
            Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1653)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .1653
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="KATO III"
                     /cell_type="signet-ring cell carcinoma"
                     /clone="KAT09742"
                     /clone_lib="KAT"
                     /note="cloning vector pME18SFL3"
BASE COUNT      441 a    369 c    358 g    485 t
ORIGIN


  Query Match           57.5%;  Score 23;  DB 37;  Length 1653;
  Best Local Similarity  74.4%;  Pred. No. 23;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db    869 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 831


RESULT   2
HSTEGT/c
LOCUS       HSTEGT     2600 bp    mRNA            PRI       28-SEP-1995
DEFINITION  H.sapiens TEGT gene.
ACCESSION   X75861
VERSION     X75861.1  GI:456258
KEYWORDS    TEGT gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2600)
  AUTHORS   Walter,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-1994) L. Walter, Abteilung Immungenetik der
            Universitaet Goettingen, Gosslerstr 12d, 37073 Goettingen, FRG
REFERENCE   2  (bases 1 to 2600)
  AUTHORS   Walter,L., Marynen,P., Szpirer,J., Levan,G. and Gunther,E.
  TITLE     Identification of a novel conserved human gene, TEGT
  JOURNAL   Genomics 28 (2), 301-304 (1995)
  MEDLINE   96015061
FEATURES             Location/Qualifiers
     source          1. .2600
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12q12-q13"
                     /clone_lib="lambda gt11 cDNA"
                     /tissue_type="testis"
     gene            41. .754
                     /gene="TEGT"
     CDS             41. .754
                     /gene="TEGT"
                     /codon_start=1
                     /protein_id="CAA53472.1"
                     /db_xref="GI:458545"
                     /db_xref="SWISS-PROT:P55061"
                     /translation="MNIFDRKINFDALLKFSHITPSTQQHLKKVYASFALCMFVAAAG
                     AYVHMVTHFIQAGLLSALGSLILMIWLMATPHSHETEQKRLGLLAGFAFLTGVGLGPA
                     LEFCIAVNPSILPTAFMGTAMIFTCFTLSALYARRRSYLFLGGILMSALSLLLLSSLG
                     NVFFGSIWPFQANLYVGLVVMCGFVLVDTQLIIEKAEHGDQDYIWHCIDLFLDFITVF
                     RKLMMILAMNEKDKKKEKK"
BASE COUNT      619 a    621 c    590 g    770 t
ORIGIN


  Query Match           57.5%;  Score 23;  DB 78;  Length 2600;
  Best Local Similarity  74.4%;  Pred. No. 23;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db   2024 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 1986


RESULT   3
AR058920/c
LOCUS       AR058920     2634 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5837838.
ACCESSION   AR058920
VERSION     AR058920.1  GI:5984497
```

```
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2634)
  AUTHORS   Reed,J.C. and Xu,Q.
  TITLE     Bax inhibitor proteins
  JOURNAL   Patent: US 5837838-A 1 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .2634
                     /organism="unknown"
BASE COUNT      632 a    626 c    597 g    779 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 81;  Length 2634;
  Best Local Similarity   74.4%;  Pred. No. 23;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db   2056 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 2018


RESULT   4
AR058921
LOCUS       AR058921     2634 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 2 from patent US 5837838.
ACCESSION   AR058921
VERSION     AR058921.1  GI:5984498
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2634)
  AUTHORS   Reed,J.C. and Xu,Q.
  TITLE     Bax inhibitor proteins
  JOURNAL   Patent: US 5837838-A 2 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .2634
                     /organism="unknown"
BASE COUNT      779 a    597 c    626 g    632 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 81;  Length 2634;
  Best Local Similarity   74.4%;  Pred. No. 23;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db    579 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 617


RESULT   5
AC019168
LOCUS       AC019168   197189 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 12 clone RP11-161E16, WORKING DRAFT
            SEQUENCE, 24 unordered pieces.
ACCESSION   AC019168
VERSION     AC019168.3  GI:7230912
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 197189)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197189)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:7021812.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0161E16
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 92%
            Sequencing vector: plasmid; 8%
            Chemistry: Dye-primer ET; 92% of reads
            Chemistry: Dye-terminator Big Dye; 8% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 180505 bases at least Q40
            Consensus quality: 186042 bases at least Q30
            Consensus quality: 189445 bases at least Q20
            Insert size: 190000; agarose-fp
            Insert size: 194889; sum-of-contigs
            Quality coverage: 3.69 in Q20 bases; agarose-fp
            Quality coverage: 3.65 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1451: contig of 1451 bp in length
            *    1452     1551: gap of unknown length
            *    1552     2906: contig of 1355 bp in length
            *    2907     3006: gap of unknown length
            *    3007     4304: contig of 1298 bp in length
            *    4305     4404: gap of unknown length
            *    4405     5621: contig of 1217 bp in length
            *    5622     5721: gap of unknown length
            *    5722     8099: contig of 2378 bp in length
            *    8100     8199: gap of unknown length
            *    8200    10800: contig of 2601 bp in length
            *   10801    10900: gap of unknown length
            *   10901    14200: contig of 3300 bp in length
            *   14201    14300: gap of unknown length
            *   14301    17571: contig of 3271 bp in length
            *   17572    17671: gap of unknown length
            *   17672    21856: contig of 4185 bp in length
            *   21857    21956: gap of unknown length
            *   21957    25748: contig of 3792 bp in length
            *   25749    25848: gap of unknown length
            *   25849    29842: contig of 3994 bp in length
            *   29843    29942: gap of unknown length
            *   29943    35421: contig of 5479 bp in length
            *   35422    35521: gap of unknown length
            *   35522    40974: contig of 5453 bp in length
            *   40975    41074: gap of unknown length
            *   41075    48591: contig of 7517 bp in length
            *   48592    48691: gap of unknown length
            *   48692    58168: contig of 9477 bp in length
            *   58169    58268: gap of unknown length
            *   58269    68032: contig of 9764 bp in length
            *   68033    68132: gap of unknown length
            *   68133    77463: contig of 9331 bp in length
            *   77464    77563: gap of unknown length
            *   77564    93029: contig of 15466 bp in length
            *   93030    93129: gap of unknown length
            *   93130   107954: contig of 14825 bp in length
            *  107955   108054: gap of unknown length
            *  108055   121842: contig of 13788 bp in length
            *  121843   121942: gap of unknown length
```

```
          *   121943   137362: contig of 15420 bp in length
          *   137363   137462: gap of unknown length
          *   137463   155524: contig of 18062 bp in length
          *   155525   155624: gap of unknown length
          *   155625   173416: contig of 17792 bp in length
          *   173417   173516: gap of unknown length
          *   173517   197189: contig of 23673 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197189
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-161E16"
     misc_feature    1. .1451
                     /note="assembly_name:Contig3"
     misc_feature    1552. .2906
                     /note="assembly_name:Contig4"
     misc_feature    3007. .4304
                     /note="assembly_name:Contig5"
     misc_feature    4405. .5621
                     /note="assembly_name:Contig6"
     misc_feature    5722. .8099
                     /note="assembly_name:Contig7"
     misc_feature    8200. .10800
                     /note="assembly_name:Contig8"
     misc_feature    10901. .14200
                     /note="assembly_name:Contig9"
     misc_feature    14301. .17571
                     /note="assembly_name:Contig10"
     misc_feature    17672. .21856
                     /note="assembly_name:Contig11"
     misc_feature    21957. .25748
                     /note="assembly_name:Contig12"
     misc_feature    25849. .29842
                     /note="assembly_name:Contig13"
     misc_feature    29943. .35421
                     /note="assembly_name:Contig14"
     misc_feature    35522. .40974
                     /note="assembly_name:Contig15"
     misc_feature    41075. .48591
                     /note="assembly_name:Contig16"
     misc_feature    48692. .58168
                     /note="assembly_name:Contig17"
     misc_feature    58269. .68032
                     /note="assembly_name:Contig18"
     misc_feature    68133. .77463
                     /note="assembly_name:Contig19"
     misc_feature    77564. .93029
                     /note="assembly_name:Contig20"
     misc_feature    93130. .107954
                     /note="assembly_name:Contig21
                     clone_end:SP6
                     vector_side:right"
     misc_feature    108055. .121842
                     /note="assembly_name:Contig22"
     misc_feature    121943. .137362
                     /note="assembly_name:Contig23"
     misc_feature    137463. .155524
                     /note="assembly_name:Contig24"
     misc_feature    155625. .173416
                     /note="assembly_name:Contig25"
     misc_feature    173517. .197189
                     /note="assembly_name:Contig26"
BASE COUNT    51312 a  47093 c  47437 g  49028 t   2319 others
ORIGIN


  Query Match             57.5%;  Score 23;  DB 42;  Length 197189;
  Best Local Similarity   74.4%;  Pred. No. 22;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||| .
```

```
Db 123952 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 123990


RESULT   6
AP002424
LOCUS       AP002424    171185 bp    DNA             HTG       31-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-873L22 map 18q21, WORKING
            DRAFT SEQUENCE, in unordered pieces.
ACCESSION   AP002424
VERSION     AP002424.1  GI:8131688
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-873L22.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171185)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 171,185 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 171185)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2000) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     ---------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            ---------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-873L22
            ---------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 156977 bases at least Q40
                Consensus quality: 164110 bases at least Q30
                Consensus quality: 167421 bases at least Q20
                Insert size: 169485; sum-of-contigs
                Quality coverage: 4.80x in Q20 bases; sum-of-contigs
            ----------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            18 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     24923 contig of  24923 bp in length
              25024     47004 contig of  21981 bp in length
              47105     67070 contig of  19966 bp in length
              67171     82757 contig of  15587 bp in length
              82858     95544 contig of  12687 bp in length
              95645    109592 contig of  13948 bp in length
             109693    120167 contig of  10475 bp in length
             120268    129255 contig of   8988 bp in length
             129356    137003 contig of   7648 bp in length
             137104    144376 contig of   7273 bp in length
             144477    150659 contig of   6183 bp in length
             150760    155896 contig of   5137 bp in length
             155997    160522 contig of   4526 bp in length
             160623    163263 contig of   2641 bp in length
             163364    165194 contig of   1831 bp in length
             165295    167704 contig of   2410 bp in length
             167805    169750 contig of   1946 bp in length
             169851    171185 contig of   1335 bp in length.
```

```
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .171185
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-873L22"
                     /map="18q21"
     misc_feature    1. .24923
                     /note="assembly_fragment"
     misc_feature    25024. .47004
                     /note="assembly_fragment"
     misc_feature    47105. .67070
                     /note="assembly_fragment"
     misc_feature    67171. .82757
                     /note="assembly_fragment"
     misc_feature    82858. .95544
                     /note="assembly_fragment"
     misc_feature    95645. .109592
                     /note="assembly_fragment"
     misc_feature    109693. .120167
                     /note="assembly_fragment"
     misc_feature    120268. .129255
                     /note="assembly_fragment"
     misc_feature    129356. .137003
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    137104. .144376
                     /note="assembly_fragment"
     misc_feature    144477. .150659
                     /note="assembly_fragment"
     misc_feature    150760. .155896
                     /note="assembly_fragment"
     misc_feature    155997. .160522
                     /note="assembly_fragment"
     misc_feature    160623. .163263
                     /note="assembly_fragment"
     misc_feature    163364. .165194
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    165295. .167704
                     /note="assembly_fragment"
     misc_feature    167805. .169750
                     /note="assembly_fragment"
     misc_feature    169851. .171185
                     /note="assembly_fragment"
BASE COUNT    52031 a  32086 c  33289 g  52078 t   1701 others
ORIGIN


  Query Match             57.0%;  Score 22.8;  DB 70;  Length 171185;
  Best Local Similarity   79.4%;  Pred. No. 27;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccaggg 34
          ||| |||| |||||| |||||| | |||||| |
Db  93406 CTCTAGCCTCCCCATGCCCCTCATTTCTCCAGTG 93439


RESULT   7
AP001569
LOCUS       AP001569   177097 bp    DNA             HTG       30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-859C21 map 18q21, WORKING
            DRAFT SEQUENCE, in unordered pieces.
ACCESSION   AP001569
VERSION     AP001569.2  GI:8117403
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-859C21.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177097)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 177,097 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 177097)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2000) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7380904.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-859C21
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 160670 bases at least Q40
                Consensus quality: 168058 bases at least Q30
                Consensus quality: 171631 bases at least Q20
                Insert size: 173997; sum-of-contigs
                Quality coverage: 4.63x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            32 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved.
                  1     23792 contig of  23792 bp in length
              23893     45465 contig of  21573 bp in length
              45566     58167 contig of  12602 bp in length
              58268     71173 contig of  12906 bp in length
              71274     81253 contig of   9980 bp in length
              81354     91080 contig of   9727 bp in length
              91181     97387 contig of   6207 bp in length
              97488    104902 contig of   7415 bp in length
             105003    111107 contig of   6105 bp in length
             111208    116851 contig of   5644 bp in length
             116952    123585 contig of   6634 bp in length
             123686    128985 contig of   5300 bp in length
             129086    134233 contig of   5148 bp in length
             134334    138588 contig of   4255 bp in length
             138689    142573 contig of   3885 bp in length
             142674    146286 contig of   3613 bp in length
             146397    150053 contig of   3667 bp in length
             150154    154046 contig of   3893 bp in length
             154147    156225 contig of   2079 bp in length
             156326    158578 contig of   2253 bp in length
             158679    161124 contig of   2446 bp in length
             161225    163559 contig of   2335 bp in length
             163660    165538 contig of   1879 bp in length
             165639    166949 contig of   1311 bp in length
             167050    168240 contig of   1191 bp in length
             168341    169579 contig of   1239 bp in length
             169680    170978 contig of   1299 bp in length
             171079    172080 contig of   1002 bp in length
             172181    173514 contig of   1334 bp in length
             173615    174614 contig of   1000 bp in length
             174715    175976 contig of   1262 bp in length
             176077    177097 contig of   1021 bp in length
            Sequence updated (26-May-2000).
```

```
          * NOTE: This is a 'working draft' sequence.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
FEATURES             Location/Qualifiers
     source          1. .177097
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-859C21"
                     /map="18q21"
     misc_feature    1. .23792
                     /note="assembly_fragment"
     misc_feature    23893. .45465
                     /note="assembly_fragment"
     misc_feature    45566. .58167
                     /note="assembly_fragment"
     misc_feature    58268. .71173
                     /note="assembly_fragment"
     misc_feature    71274. .81253
                     /note="assembly_fragment"
     misc_feature    81354. .91080
                     /note="assembly_fragment"
     misc_feature    91181. .97387
                     /note="assembly_fragment"
     misc_feature    97488. .104902
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    105003. .111107
                     /note="assembly_fragment"
     misc_feature    111208. .116851
                     /note="assembly_fragment"
     misc_feature    116952. .123585
                     /note="assembly_fragment"
     misc_feature    123686. .128985
                     /note="assembly_fragment"
     misc_feature    129086. .134233
                     /note="assembly_fragment"
     misc_feature    134334. .138588
                     /note="assembly_fragment"
     misc_feature    138689. .142573
                     /note="assembly_fragment"
     misc_feature    142674. .146286
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    146387. .150053
                     /note="assembly_fragment"
     misc_feature    150154. .154046
                     /note="assembly_fragment"
     misc_feature    154147. .156225
                     /note="assembly_fragment"
     misc_feature    156326. .158578
                     /note="assembly_fragment"
     misc_feature    158679. .161124
                     /note="assembly_fragment"
     misc_feature    161225. .163559
                     /note="assembly_fragment"
     misc_feature    163660. .165538
                     /note="assembly_fragment"
     misc_feature    165639. .166949
                     /note="assembly_fragment"
     misc_feature    167050. .168240
                     /note="assembly_fragment"
     misc_feature    168341. .169579
                     /note="assembly_fragment"
     misc_feature    169680. .170978
                     /note="assembly_fragment"
     misc_feature    171079. .172080
                     /note="assembly_fragment"
     misc_feature    172181. .173514
                     /note="assembly_fragment"
     misc_feature    173615. .174614
                     /note="assembly_fragment"
     misc_feature    174715. .175976
                     /note="assembly_fragment"
     misc_feature    176077. .177097
                     /note="assembly_fragment"
BASE COUNT    53531 a  33044 c  34236 g  53184 t   3102 others
ORIGIN


  Query Match          57.0%;  Score 22.8;  DB 70;  Length 177097;
  Best Local Similarity   79.4%;  Pred. No. 27;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 ctcgagccacccccataaccctcaatactccaggg 34
           ||| |||| |||||| |||||| | |||||| |
Db    3709 CTCTAGCCTCCCCATGCCCCTCATTTCTCCAGTG 3742


RESULT   8
AC007052/c
LOCUS       AC007052;  179726 bp    DNA             PRI       23-MAR-1999
DEFINITION  Homo sapiens chromosome 18, clone hRPK.411_H_24, complete sequence.
ACCESSION   AC007052
VERSION     AC007052.4  GI:4510438
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179726)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone hRPK.411_H_24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179726)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 179726)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 24, 1999 this sequence version replaced gi:4432872.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
```

source	1. .179726
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="hRPK.411_H_24"
	/clone_lib="RPCI-11 human BAC library"
	/map="18"
	/chromosome="18"
repeat_region	898. .1014
	/rpt_family="L1MD3"
repeat_region	complement(1035. .1333)
	/rpt_family="AluSx"
repeat_region	2987. .3030
	/rpt_family="CT-rich"
repeat_region	3087. .3291
	/rpt_family="L1MC4"
repeat_region	4298. .4330
	/rpt_family="(CA)n"
repeat_region	5344. .5365
	/rpt_family="AT_rich"
repeat_region	complement(5552. .5724)
	/rpt_family="AluJ"
repeat_region	5731. .11847
	/rpt_family="L1PA2"
repeat_region	11857. .11891
	/rpt_family="(TAAA)n"
repeat_region	complement(11893. .12002)
	/rpt_family="AluJ"
unsure	11925. .11948
	/note="Single-stranded coverage."
repeat_region	12026. .12321
	/rpt_family="AluSg"
repeat_region	complement(12605. .12846)
	/rpt_family="MIR"
repeat_region	13352. .13660
	/rpt_family="AluSc"
repeat_region	complement(13741. .14128)
	/rpt_family="MLT1B"
repeat_region	14425. .14456
	/rpt_family="AT_rich"
repeat_region	14457. .14978
	/rpt_family="MLT1F"
repeat_region	15496. .15834
	/rpt_family="THE1B"
repeat_region	16325. .16628
	/rpt_family="AluYb8"
repeat_region	16753. .17086
	/rpt_family="MER7A"
repeat_region	17543. .17601
	/rpt_family="MIR"
repeat_region	18075. .18442
	/rpt_family="MLT1B"
repeat_region	complement(18735. .18882)
	/rpt_family="MIR"
repeat_region	complement(18948. .19083)
	/rpt_family="L1MB8"
repeat_region	complement(19190. .19426)
	/rpt_family="AluJo"
repeat_region	19431. .19750
	/rpt_family="L1ME3"
repeat_region	20434. .20555
	/rpt_family="MLT1A1"
repeat_region	20556. .20863
	/rpt_family="AluSx"
repeat_region	20864. .21080
	/rpt_family="MLT1A1"
repeat_region	21528. .21565
	/rpt_family="AT_rich"
repeat_region	21575. .21668
	/rpt_family="L2"
repeat_region	complement(21849. .21991)
	/rpt_family="MER5B"
repeat_region	23147. .23168
	/rpt_family="(TTTG)n"
repeat_region	complement(23169. .23384)
	/rpt_family="AluJo"
repeat_region	23385. .23419
	/rpt_family="AT_rich"
repeat_region	complement(23799. .23846)
	/rpt_family="L2"
repeat_region	complement(23947. .24060)
	/rpt_family="L2"
repeat_region	24066. .24206
	/rpt_family="BC200"
repeat_region	complement(24347. .24514)
	/rpt_family="MER21B"
repeat_region	24782. .25203
	/rpt_family="L1MB2"
repeat_region	25385. .26005
	/rpt_family="L1MB2"
repeat_region	26035. .26336
	/rpt_family="AluSx"
repeat_region	26432. .26604
	/rpt_family="L1ME"
repeat_region	26610. .26645
	/rpt_family="AT_rich"
repeat_region	complement(27113. .27422)
	/rpt_family="AluSp"
repeat_region	27865. .28190
	/rpt_family="AluSg"
repeat_region	28250. .28272
	/rpt_family="AT_rich"
repeat_region	28363. .28496
	/rpt_family="HAL1"
repeat_region	complement(28497. .28796)
	/rpt_family="AluSg"
repeat_region	28797. .28869
	/rpt_family="HAL1"
repeat_region	28993. .29027
	/rpt_family="AT_rich"
repeat_region	29107. .29134
	/rpt_family="(TATATG)n"
repeat_region	29151. .29182
	/rpt_family="(TG)n"
repeat_region	complement(29821. .30164)
	/rpt_family="MER7A"
repeat_region	complement(30236. .30522)
	/rpt_family="AluJo"
repeat_region	30576. .30723
	/rpt_family="MIR"
repeat_region	30927. .31271
	/rpt_family="MLT1K"
repeat_region	32382. .32437
	/rpt_family="(CA)n"
repeat_region	32480. .32502
	/rpt_family="(TTTG)n"
repeat_region	complement(32503. .32751)
	/rpt_family="AluSx"
repeat_region	34312. .34476
	/rpt_family="MIR"
repeat_region	complement(34648. .34953)
	/rpt_family="AluSx"
repeat_region	complement(35550. .35947)
	/rpt_family="THE1B"
repeat_region	36048. .36070
	/rpt_family="(TTTTA)n"
repeat_region	complement(36072. .36366)
	/rpt_family="AluJo"
repeat_region	36525. .36587
	/rpt_family="AT_rich"
repeat_region	36792. .36849
	/rpt_family="AT_rich"
repeat_region	complement(36903. .37092)
	/rpt_family="L1MA7"
repeat_region	37502. .37812

Query Match 57.0%; Score 22.8; DB 9; Length 179726;

```
  Best Local Similarity   79.4%;  Pred. No. 27;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 ctcgagccaccccataaccctcaatactccaggg 34
           ||| |||| ||||||  |||||| | |||||| |
Db   78718 CTCTAGCCTCCCCATGCCCCTCATTTCTCCAGTG 78685


RESULT   9
AL159996
LOCUS       AL159996   188357 bp    DNA             PRI       24-JUL-2000
DEFINITION  Human DNA sequence from clone RP11-401F12 on chromosome 9. Contains
            STSs and GSSs, complete sequence.
ACCESSION   AL159996
VERSION     AL159996.7  GI:8648450
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188357)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 21, 2000 this sequence version replaced gi:8346239.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-401F12 is from the library RPCI-11.2 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-401F12 The true
            left end of clone RP11-208G24 is at 167549 in this sequence. The
            true right end of clone RP11-395D3 is at 5200 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .188357
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-401F12"
                     /clone_lib="RPCI-11.2"
     misc_feature    410. .897
                     /note="match: GSS: Em:AQ606829"
     misc_feature    1734. .2348
                     /note="match: GSS: Em:AQ823967"
     repeat_region   1819. .2460
                     /note="L1MC4 repeat: matches 6462. .7077 of consensus"
     repeat_region   2551. .2614
                     /note="2 copies 32 mer 96% conserved"
     misc_feature    2690. .2698
                     /note="TN10 excised: This sequence represents the
                     duplicated flanking sequence of the TN10."
repeat_region   3529. .3560
                /note="8 copies 4 mer gtgt 96% conserved"
repeat_region   3746. .3959
                /note="L1PB3 repeat: matches 5916. .6147 of consensus"
repeat_region   3969. .4006
                /note="U2 repeat: matches 1. .38 of consensus"
repeat_region   4392. .4430
                /note="13 copies 3 mer aac 92% conserved"
misc_feature    5217. .5472
                /note="match: GSS: Em:AQ302480"
misc_feature    5260. .6024
                /note="match: GSS: Em:AQ901451"
misc_feature    complement(5743. .6091)
                /note="match: GSS: Em:AQ132828"
misc_feature    6101. .6658
                /note="match: GSS: Em:AQ830958"
repeat_region   7246. .8450
                /note="Tigger3b repeat: matches 8. .1231 of consensus"
repeat_region   8626. .8773
                /note="L2 repeat: matches 2452. .2611 of consensus"
repeat_region   10391. .10584
                /note="MER5A repeat: matches 1. .189 of consensus"
repeat_region   11880. .12174
                /note="L2 repeat: matches 1986. .2272 of consensus"
misc_feature    complement(12504. .12890)
                /note="match: STS: Em:HSA311YC1"
repeat_region   12609. .12648
                /note="10 copies 4 mer gtgt 90% conserved"
repeat_region   12610. .12647
                /note="19 copies 2 mer tg 92% conserved"
repeat_region   12615. .12650
                /note="6 copies 6 mer gtgtgt 91% conserved"
repeat_region   13805. .14099
                /note="AluSg repeat: matches 2. .295 of consensus"
repeat_region   14101. .14388
                /note="AluSq repeat: matches 16. .302 of consensus"
repeat_region   15605. .15839
                /note="L2 repeat: matches 2357. .2612 of consensus"
repeat_region   15947. .16013
                /note="L2 repeat: matches 2609. .2674 of consensus"
repeat_region   16084. .16364
                /note="AluJo repeat: matches 1. .288 of consensus"
misc_feature    complement(16254. .16783)
                /note="match: GSS: Em:AQ764225"
repeat_region   17436. .17528
                /note="MIR repeat: matches 35. .123 of consensus"
repeat_region   18177. .18298
                /note="FLAM_A repeat: matches 1. .122 of consensus"
repeat_region   19380. .19675
                /note="AluSx repeat: matches 1. .296 of consensus"
misc_feature    complement(19534. .19956)
                /note="match: GSS: Em:AQ042003"
repeat_region   19736. .19958
                /note="L2 repeat: matches 2477. .2745 of consensus"
repeat_region   20126. .20237
                /note="L2 repeat: matches 2589. .2705 of consensus"
repeat_region   20737. .20812
                /note="MADE1 repeat: matches 5. .80 of consensus"
repeat_region   21025. .21200
                /note="L2 repeat: matches 2184. .2355 of consensus"
repeat_region   21443. .21646
                /note="MIR repeat: matches 40. .246 of consensus"
repeat_region   23474. .23793
                /note="MLT1F repeat: matches 197. .541 of consensus"
misc_feature    complement(25531. .26230)
                /note="match: GSS: Em:AQ787911"
repeat_region   27777. .28867
                /note="L1MC3 repeat: matches 6667. .7734 of consensus"
repeat_region   28802. .28880
                /note="L1MC2 repeat: matches 6247. .6326 of consensus"
repeat_region   28924. .29133
                /note="MER20 repeat: matches 2. .214 of consensus"
repeat_region   29810. .29849
```

```
                     /note="20 copies 2 mer ac 80% conserved"
     repeat_region   29868. .30157
                     /note="AluY repeat: matches 3. .290 of consensus"
     repeat_region   30822. .31276
                     /note="MLT1H repeat: matches 70. .532 of consensus"
     repeat_region   31665. .32020
                     /note="THE1B repeat: matches 1. .364 of consensus"
     misc_feature    32181. .32674
                     /note="match: GSS: Em:AQ822292"
     repeat_region   32279. .32403
                     /note="MIR repeat: matches 57. .192 of consensus"
     misc_feature    complement(32857. .33263)
                     /note="match: GSS: Em:AQ139531"
     repeat_region   33279. .33551
                     /note="AluSq repeat: matches 1. .301 of consensus"
     repeat_region   35349. .35388
                     /note="10 copies 4 mer acac 92% conserved"
     repeat_region   35353. .35388
                     /note="3 copies 12 mer 94% conserved"
     repeat_region   35759. .36002
                     /note="MIR repeat: matches 2. .246 of consensus"
     misc_feature    36104. .36469
                     /note="match: GSS: Em:AQ031409"
     repeat_region   37497. .37545
                     /note="MIR repeat: matches 81. .127 of consensus"
     repeat_region   37664. .38248
                     /note="MLT2D repeat: matches 1. .553 of consensus"
     repeat_region   38259. .38373
                     /note="MIR repeat: matches 122. .249 of consensus"
     repeat_region   38334. .38392
                     /note="L2 repeat: matches 2692. .2750 of consensus"
     misc_feature    38609. .38887
                     /note="match: STS: Em:G13414"
     repeat_region   38781. .38991
                     /note="LTR33 repeat: matches 302. .518 of consensus"
     repeat_region   39086. .39212
                     /note="LTR33 repeat: matches 29. .165 of consensus"
     repeat_region   39249. .39483
                     /note="MER4D repeat: matches 1. .234 of consensus"
     repeat_region   39483. .40142
                     /note="MER4D repeat: matches 359. .1017 of consensus"
     repeat_region   40325. .40415
                     /note="MER94 repeat: matches 39. .134 of consensus"
     repeat_region   41795. .41902
                     /note="LTR33 repeat: matches 370. .472 of consensus"
     repeat_region   41903. .42192
                     /note="AluSg repeat: matches 1. .291 of consensus"
     repeat_region   42193. .42241
                     /note="LTR33 repeat: matches 472. .521 of consensus"
     repeat_region   43494. .43613
                     /note="L2 repeat: matches 2628. .2748 of consensus"
     repeat_region   43695. .43862
                     /note="FRAM repeat: matches 0. .164 of consensus"
     repeat_region   44019. .44082
                     /note="L2 repeat: matches 2356. .2428 of consensus"
     repeat_region   44083. .44384
                     /note="L1PA4 repeat: matches 5841. .6144 of consensus"
     repeat_region   44385. .44529
                     /note="L2 repeat: matches 2252. .2356 of consensus"
     repeat_region   44541. .44628
                     /note="MIR repeat: matches 128. .229 of consensus"
     repeat_region   44631. .44747
                     /note="MER33 repeat: matches 203. .324 of consensus"
     repeat_region   44779. .44985

  Query Match             57.0%;  Score 22.8;  DB 37;  Length 188357;
  Best Local Similarity   79.4%;  Pred. No. 27;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       7 ccaccccataaccctcaatactccagggattggg 40
           |||||||||  | ||||| |||||||||   |||
Db   56526 CCACCCCATGTCTCTCAAGACTCCAGGGGAAGGG 56559
```

```
RESULT  10
AP001592
LOCUS       AP001592   200774 bp    DNA             HTG       29-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-869L2 map 18q21, WORKING
            DRAFT SEQUENCE, in unordered pieces.
ACCESSION   AP001592
VERSION     AP001592.3  GI:9588073
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-869L2.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200774)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 200,774 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 200774)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitasato,
            Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Jul 31, 2000 this sequence version replaced gi:8117416.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-869L2
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 197367 bases at least Q40
                Consensus quality: 198735 bases at least Q30
                Consensus quality: 199244 bases at least Q20
                Insert size: 199774; sum-of-contigs
                Quality coverage: 9.59x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            11 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                    1     36889 contig of  36889 bp in length
                36990     73347 contig of  36358 bp in length
                73448     98904 contig of  25457 bp in length
                99005    124595 contig of  25591 bp in length
               124696    145130 contig of  20435 bp in length
               145231    163732 contig of  18502 bp in length
               163833    181811 contig of  17979 bp in length
               181912    190895 contig of   8984 bp in length
               190996    198298 contig of   7303 bp in length
               198399    199648 contig of   1250 bp in length
               199749    200774 contig of   1026 bp in length.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES                Location/Qualifiers
     source             1. .200774
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
```

```
                     /chromosome="18"
                     /clone="RP11-869L2"
                     /map="18q21"
     misc_feature    1. .36889
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    36990. .73347
                     /note="assembly_fragment"
     misc_feature    73448. .98904
                     /note="assembly_fragment"
     misc_feature    99005. .124595
                     /note="assembly_fragment"
     misc_feature    124696. .145130
                     /note="assembly_fragment"
     misc_feature    145231. .163732
                     /note="assembly_fragment"
     misc_feature    163833. .181811
                     /note="assembly_fragment"
     misc_feature    181912. .190895
                     /note="assembly_fragment"
     misc_feature    190996. .198298
                     /note="assembly_fragment"
     misc_feature    198399. .199648
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    199749. .200774
                     /note="assembly_fragment"
BASE COUNT    61549 a  39812 c  38979 g  59434 t   1000 others
ORIGIN


  Query Match             57.0%;  Score 22.8;  DB 70;  Length 200774;
  Best Local Similarity   79.4%;  Pred. No. 27;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 ctcgagccaccccataaccctcaatactccaggg 34
           ||| |||| |||||  |||||| | |||||| |
Db   47483 CTCTAGCCTCCCCATGCCCCTCATTTCTCCAGTG 47516


RESULT  11
HS471M13
LOCUS       HS471M13   151514 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from PAC 471M13 on chromosome 1q24. Contains
            ESTs and STS.
ACCESSION   Z97198
VERSION     Z97198.1  GI:3169111
KEYWORDS    1q24.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151514)
  AUTHORS   Patel,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1997) Chromosome 1 Project Group
            (http://www.sanger.ac.uk/HGP/Chr1/) Sanger Centre, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 30, 1998 this sequence version replaced gi:2225929.
            IMPORTANT: This sequence is the entire insert of clone 471M13.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre chromosome 1
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 471M13 is at 1 in this sequence. The
            true right end of clone 436E22 is at 37386.
            The true right end of clone 332O11 is at 50278.
            The true left end of clone 436E22 is at 148355.
            The true right end of clone 471M13 is at 151514.
            471M13 is from the library RPCI3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES             Location/Qualifiers
     source          1. .151514
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q24"
                     /clone="RP3-471M13"
                     /clone_lib="RPCI-3"
     repeat_region   3125. .3421
                     /note="AluJb repeat: matches 300. .2 of consensus"
     repeat_region   3679. .3832
                     /note="MLT1E repeat: matches 568. .409 of consensus"
     repeat_region   3835. .3920
                     /note="MLT1E repeat: matches 379. .295 of consensus"
     repeat_region   4053. .4177
                     /note="MLT1D repeat: matches 241. .116 of consensus"
     repeat_region   5801. .5858
                     /note="L1MA2 repeat: matches 1055. .998 of consensus"
     repeat_region   6310. .6538
                     /note="MER20 repeat: matches 1. .218 of consensus"
     repeat_region   6786. .7168
                     /note="MLT1B repeat: matches 390. .2 of consensus"
     repeat_region   9614. .9761
                     /note="MIR2 repeat: matches 145. .1 of consensus"
     repeat_region   12836. .12884
                     /note="MARINER repeat: matches 1224. .1276 of consensus"
     repeat_region   12847. .12883
                     /note="MADE1 repeat: matches 37. .1 of consensus"
     repeat_region   14595. .14636
                     /note="21 copies of 2 mer 100 % conserved"
     repeat_region   14653. .14780
                     /note="MIR repeat: matches 57. .192 of consensus"
     repeat_region   16389. .16438
                     /note="25 copies of 2 mer 94 % conserved"
     repeat_region   18276. .19016
                     /note="L1MB4 repeat: matches 737. .1 of consensus"
     repeat_region   18866. .22120
                     /note="L1 repeat: matches 5390. .2079 of consensus"
     repeat_region   22440. .22637
                     /note="AluY repeat: matches 95. .291 of consensus;
                     incomplete repeat"
     repeat_region   22820. .23441
                     /note="L1 repeat: matches 1986. .1359 of consensus"
     repeat_region   23172. .23612
                     /note="MER25 repeat: matches 2136. .1704 of consensus"
     repeat_region   23900. .24186
                     /note="MER43 repeat: matches 2. .272 of consensus"
     repeat_region   25649. .25801
                     /note="MIR repeat: matches 77. .236 of consensus"
     repeat_region   26135. .26308
                     /note="MER45 repeat: matches 4. .177 of consensus"
     repeat_region   27338. .27424
                     /note="MIR repeat: matches 60. .146 of consensus"
     repeat_region   27652. .28554
                     /note="L1MB7 repeat: matches 920. .1 of consensus"
     repeat_region   28415. .29325
                     /note="L1 repeat: matches 5390. .4467 of consensus"
     repeat_region   29328. .29630
                     /note="AluSx repeat: matches 1. .296 of consensus"
     repeat_region   29708. .30974
                     /note="L1 repeat: matches 4402. .3062 of consensus"
     repeat_region   31008. .31171
                     /note="AluJo repeat: matches 295. .122 of consensus;
```

```
                incomplete repeat"
repeat_region   31419. .31921
                /note="MER21B repeat: matches 790. .280 of consensus"
prim_transcript 36489. .37297
                /note="match: 5' EST T59288 clone 79506; Paired with EST
                T59247 matching this clone; match: 3' EST T59247 clone
                79506; Paired with EST T59288 matching this clone"
prim_transcript <37387. .>37782
                /note="match: 3' EST H94492 clone 243199"
repeat_region   38289. .38376
                /note="4 copies of 22 mer 84 % conserved"
repeat_region   39657. .39904
                /note="pTR5 repeat: matches 505. .751 of consensus"
repeat_region   39925. .40627
                /note="pTR5 repeat: matches 1693. .2427 of consensus"
repeat_region   40860. .41154
                /note="AluSx repeat: matches 296. .1 of consensus"
repeat_region   41454. .41745
                /note="AluSx repeat: matches 2. .295 of consensus"
repeat_region   42718. .43610
                /note="L1PA2 repeat: matches 893. .1 of consensus"
repeat_region   43461. .44234
                /note="L1 repeat: matches 5390. .4619 of consensus"
variation       43947. .43949
                /note="clone 346E22; GAA in this entry; substitution"
                /replace="gca"
variation       44028. .44030
                /note="clone 346E22; CGG in this entry; substitution"
                /replace="cag"
variation       44068. .44070
                /note="clone 346E22; ATA in this entry; substitution"
                /replace="aca"
repeat_region   44249. .45146
                /note="L1PA13 repeat: matches 903. .1 of consensus"
variation       44447. .44449
                /note="clone 346E22; CTT in this entry; substitution"
                /replace="cct"
variation       44552. .44554
                /note="clone 346E22; CGT in this entry; substitution"
                /replace="cat"
variation       44706. .44708
                /note="clone 346E22; CGC in this entry; substitution"
                /replace="cac"
repeat_region   44994. .46864
                /note="L1 repeat: matches 5389. .3510 of consensus"
repeat_region   49047. .49319
                /note="MER21B repeat: matches 320. .41 of consensus"
variation       49176. .49178
                /note="clone 346E22; GTT in this entry; substitution"
                /replace="gat"
repeat_region   49544. .49929
                /note="THE1C repeat: matches 371. .1 of consensus"
repeat_region   49930. .51453
                /note="THE1B-INTERNAL repeat: matches 1580. .1 of
                consensus"
variation       50448. .50450
                /note="clone 346E22; AGT in this entry; substitution"
                /replace="aat"
variation       50522. .50525
                /note="clone 346E22; GGTC in this entry; deletion"
                /replace="ggctc"
repeat_region   51454. .51755
                /note="THE1C repeat: matches 371. .75 of consensus"
variation       51488. .51490
                /note="clone 346E22; ATA in this entry; substitution"
                /replace="aca"
repeat_region   51936. .52202
                /note="L1 repeat: matches 1968. .1703 of consensus"
repeat_region   52980. .53015
                /note="18 copies of 2 mer 81 % conserved"
repeat_region   53822. .54122
                /note="AluJo repeat: matches 1. .300 of consensus"
repeat_region   54908. .54951
                /note="2 copies of 22 mer 100 % conserved"
repeat_region   56468. .56599
                /note="MIR repeat: matches 2. .131 of consensus"
repeat_region   57073. .57206
                /note="MIR2 repeat: matches 1. .134 of consensus"
repeat_region   58539. .58839
                /note="AluY repeat: matches 1. .301 of consensus"
repeat_region   60836. .61280
                /note="MER28 repeat: matches 434. .1 of consensus"
unsure          63474. .63782
repeat_region   65291. .65352
                /note="MSTD repeat: matches 394. .333 of consensus"
repeat_region   65446. .66331
                /note="L1PA2 repeat: matches 891. .1 of consensus"
repeat_region   66182. .66388
                /note="L1 repeat: matches 5390. .5185 of consensus"
repeat_region   66390. .66707
                /note="MSTD repeat: matches 336. .9 of consensus"
repeat_region   68956. .69185
                /note="MIR repeat: matches 5. .262 of consensus"
repeat_region   69855. .69951

  Query Match          56.0%;  Score 22.4;  DB 77;  Length 151514;
  Best Local Similarity  72.5%;  Pred. No. 40;
  Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
            ||| | || |||| | ||||||| | ||| |||| || |
Db 114008 CTCCACCCTCCCCCCATCCCTCAACAGTCCCGGGAGTGTG 114047


RESULT  12
G22123
LOCUS       G22123         452 bp    DNA             STS       31-MAY-1996
DEFINITION  human STS WI-15564, sequence tagged site.
ACCESSION   G22123
VERSION     G22123.1  GI:1342449
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: ACTATTGGGCGAAGTATCTTATCTC
            Primer B: GGTTCCGTGGTTTAAGAGTATAGA
            STS size: 138
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul
```

```
          Buffer:
              MgCl2: 1.5 mM
              KCl: 50 mM
              Tris-HCL: 10 mM
              pH: 9.3

          Derived from dbEST (genbank accession R40684).
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="295.0 cR from top of Chr19 linkage group"
     STS             31. .168
     primer_bind     31. .55
     primer_bind     complement(145. .168)
BASE COUNT      126 a     82 c    106 g    136 t      2 others
ORIGIN


  Query Match             55.0%;  Score 22;  DB 79;  Length 452;
  Best Local Similarity   73.7%;  Pred. No. 62;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       3 cgagccaccccataaccctcaatactccagggattggg 40
           ||||  ||||||||||||||   ||| ||| ||   |||
Db     361 CGAGTAACCCCATAACCCTGTTTACCCCAAGGGCAGGG 398


RESULT 13
TBR012199/c
LOCUS       TBR012199    34887 bp    DNA             INV       21-OCT-1999
DEFINITION  Trypanosoma brucei GRESAG4 genes, ESAG9 gene, ESAG9 pseudogene,
            ESAG1 pseudogene and ILTat 1.61 metacyclic VSG genes, stock EATRO
            795.
ACCESSION   AJ012199
VERSION     AJ012199.2  GI:4456876
KEYWORDS    adenylate cyclase; ESAG1 gene; ESAG9 gene; ESAG9 pseudogene;
            GRESAG4 gene; iltat 1.61 metacyclic VSG gene.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 34887)
  AUTHORS   Matthews,K.R., Shiels,P.G., Graham,S.V., Cowan,C. and Barry,J.D.
  TITLE     Duplicative activation mechanisms of two trypanosome telomeric VSG
            genes with structurally simple 5' flanks
  JOURNAL   Nucleic Acids Res. 18 (24), 7219-7227 (1990)
  MEDLINE   91081307
REFERENCE   2  (bases 1 to 34887)
  AUTHORS   Graham,S.V., Terry,S. and Barry,J.D.
  TITLE     A structural and transcription pattern for variant surface
            glycoprotein gene expression sites used in metacyclic stage
            Trypanosoma brucei
  JOURNAL   Mol. Biochem. Parasitol. 103 (2), 141-154 (1999)
  MEDLINE   20017521
REFERENCE   3  (bases 1 to 34887)
  AUTHORS   Barry,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) Barry J.D., Wellcome Unit of Molecular
            Parasitology, University of Glasgow, The Anderson College, 56
            Dumbarton Rd, Glasgow, G11 6NU, SCOTLAND
  REMARK    Revised by [4]
REFERENCE   4  (bases 1 to 34887)
  AUTHORS   Barry,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) Barry J.D., Wellcome Unit of Molecular
            Parasitology, University of Glasgow, The Anderson College, 56
            Dumbarton Rd, Glasgow, G11 6NU, SCOTLAND
COMMENT     On Mar 21, 1999 this sequence version replaced gi:4151108.
            Related sequences: X54732.
FEATURES             Location/Qualifiers
     source          1. .34887
                     /organism="Trypanosoma brucei"
                     /db_xref="taxon:5691"
                     /note="stock EATRO 795"
     gene            702. .10716
                     /gene="GRESAG4"
     CDS             702. .3989
                     /gene="GRESAG4"
                     /note="Putative surface receptor family with intracellular
                     adenylate cyclase domain"
                     /codon_start=1
                     /product="adenylate cyclase"
                     /protein_id="CAA09951.1"
                     /db_xref="GI:4151109"
                     /db_xref="SPTREMBL:O97349"
                     /translation="MYDMKADVEIVNSLNAGLSASLATNSVEKASDFEISLQTPSSYN
                     QTIEETFNDAVKNSKEQLLVIVGPLGNQNVLWIRDKLEEHDLVAFAPLSYSDEVRGWN
                     PHFYYPSVEPKAELLALIRYVVVFLRLRRVGFMYLKGTNFGDSSYTFTEQVMSMMGYK
                     LCGTFAIDGNVNERIGDDAFDREWEKFVKTRPQAVLLFGSPHDTTKKFIQYLVREDYT
                     ADAFVLTTSSSQTFLLNTWKEALEETNTELKPGQLIITGTGPLASDTRYKIIQRFREE
                     MTKYLNTNENWGGFAKPEHFDTNHNSGELMVLGWLAGEMLMRALVDSISLTNRTLFRL
                     SLYNQRRYLIDDMVIGDFGGECGSAAQLQGAVCRCNQGGSVVYMKEVVEKYQFASLRR
                     GFFLQKGSECGTLSVPLYAPMSSSIIMMADMENAHRANLRYYRGIFATFLRIFGENRD
                     RLYLHSLNSSYSGAAKDLNDTLDERIVSAVFGVVTNEVLTTPGVIFFDPMVVFPQVST
                     FKRNVIHFFPTLTQELYVLAQYFSRLTNHIAHTIIRSDEAAEVAEVLSMSLMTFGASL
                     GSRVLLNEHVADIHMEHLKHGGDVFILGLTTPDVKTISKYLDTHPQAGVFITFTDLLL
                     YYEELKLEFSVTANAKAQRLLFATSLPHWADNSSTSETIIAYHEAVPDPSQWSPMSLR
                     GFAIARVMQTLLLPMKKVSSSLLSDEIFWQTSFVVDDMRYGPFSDSDCIANGVQLSGN
                     CVWNYGATDISVWSFGRVLDPTLPVMQEPVTPSIVYKKIRVQLSSSQVVGIIIGLVFA
                     LLLFVTLGVALYCTISNKRDNELAPRQPTDPVTLIFTDIESSTAQWATYPDIMTEAVA
                     AHHRMIRQLVLKHDCYEVKTIGDSFMIACQDPFKAVQLAADLQLMFLHNDWGTDALDN
                     FYREFEEANAKEDNEYTSPTACLDSEVYCRLWNGLRVRIGIHTGLCDIRHDEVTKGFD
                     YYGPTTNMAARTESVANGGQVLLTHATYMSLSESQRHQLDVTALGPVQLRGVPQPVQL
                     YQLNAVPGRTFAALRLDHDYYFEDGNETTNSTSENSSSRVELSESAQMIMSSPSNVTQ
                     HIQRSAT"
     CDS             7426. .10716
                     /gene="GRESAG4"
                     /note="Putative surface receptor family with intracellular
                     adenylate cyclase domain"
                     /codon_start=1
                     /product="adenylate cyclase"
                     /protein_id="CAA09952.1"
                     /db_xref="GI:4151110"
                     /db_xref="SPTREMBL:O97350"
                     /translation="MYGAGFPHNVVDSLNAGLSASIAAREEKLRTLVKVSLIQLTSYE
                     ESVEQLFNTTVERTNNELLVIVGPLGNQNVLWIRDKLEEHDLVAFAPLSYFDKTRGWN
                     PHLYYPSVEPKAELLALIRYAVVFLRLRRVGFMYLKGTNFGDSSYTFTEQVMSMMGYK
                     LCGTFAISGNVNERIGDDAFDREWEKFVKTRPQAVLLFGSPHDTTKKFIQYLVTEDHH
                     TAHAFVLTTSSSQTFLLSTWKEALEETNTELKPGQLIITGTGPLASDTRYKIIQRFRE
                     EMTKYLNTNENWGGFAKPEHFDTNHNSGELMVLGWLAGEMLMRALVDSISLTNRTLFR
                     LSLYNQRRYLIDDMVIGDFGGECGSAAQLQGAVCRCNQGGSVVYMKEVVEKYHLEAVE
                     EGLLTWGTTRCFSTGIRVYAPLSGLVILISDNYVVQRANLLYFSGAVTLTGNDHIGDN
                     DRLFFHSLNSSYSGAAKDLNDTLDERIVSAVFGVLTNEVLTTPNVIFFDPMVVFPQVS
                     TFKRNVIHFFPTLTQELYVLAQYFSRLTNHIAHTIIRSDEAAEVAEVLSMSLMTFGAS
                     LGSRVLLNEHVSDIHMEHLKHGGDVFILGLTTPDVKTISKYLDTHPQARVFITFTDLL
                     LYYEELKLEFSVTANAKAQRLLFATSLPHWADNSSTSETIIAYHEAIPDPSQWSPMSL
                     RGFAIARVMQTLLLPMKKVSSSLLSDEIFWQTSFVVDDMRYGPFSDSDCIANGVQLSG
                     NCVWNYGATDISVWSFGRVLDPTLPVMQEPVTPSIVYKKIRVQLSSSQVVGIIIGLVF
                     ALLLFVTLGVALYCTISNKRDNELAPRQPTDPVTLIFTDIESSTAQWATYPDIMTEAV
                     AAHHRMIRQLVLKHDCYEVKTIGDSFMIACQDPFKAVQLAADLQLMFLHNDWGTDALD
                     NFYREFEEANAKEDNENTSPTACLDSEVYCRLWNGLRVRIGIHTGLCDIRHDEVTKGF
                     DYYGPTTNMAARTESVANGGQVLLTHATYMSLSESQRHQLDVTALGPVQLRGVPQPVQ
                     LYQLNAVPGRTFAALRLDHDYYFEDGNETTNSTSENSSSRVELSESAQMIMSSPSNVT
                     QHIQRSAT"
     gene            15148. .22722
                     /gene="ESAG9, putative"
     CDS             <15148. .15939
                     /gene="ESAG9, putative"
                     /codon_start=1
                     /protein_id="CAA09953.1"
                     /db_xref="GI:4151111"
                     /db_xref="SPTREMBL:O97351"
                     /translation="MFENATVLLLTLITCLLPAESRSASATTSVVGLAVSSGSGCIQY
                     WDRDGRGVPCSSLPSRRGRPSGHTSPQLPVLQPQTQRREVRGSAIGAQHGGRRVPSPA
```

```
                     LSISSGTRSGSLGAPQEGQPQSDAASATGIGGQVERNDGPAGLSSREEPATTEVEESR
                     QATVRVKEVSESEKRDRGSESHEWQKQGRNTTPPPLTRESGEGLAHPQPRSAVSEPDQ
                     GSKQEAATSDLEEHESTENSTQGRRSVAHTRNAMILSAVLCLLSY"
     CDS             <21835. .22722
                     /gene="ESAG9, putative"
                     /codon_start=1
                     /pseudo
     gene            26442. .27427
                     /gene="ESAG1"
     CDS             26442. .27427
                     /gene="ESAG1"
                     /note="Internal stop codons & frame shifts"
                     /codon_start=1
                     /pseudo
     promoter        31036
                     /gene="VSG-1.61"
                     /note="Transcription start site observed in metacyclic
                     trypanosome clones"
                     /evidence=experimental
     gene            31036
                     /gene="VSG-1.61"
     repeat_region   31671. .31832
                     /citation=[1]
                     /rpt_family="70-bp"
                     /rpt_unit=31671. .31747
                     /rpt_type=DIRECT
     gene            33188. .34744
                     /gene="ILTat 1.61 metacyclic VSG"
     CDS             33188. .34744
                     /gene="ILTat 1.61 metacyclic VSG"
                     /codon_start=1
                     /protein_id="CAA09956.1"
                     /db_xref="GI:4151112"
                     /db_xref="SPTREMBL:O97352"
                     /translation="MDLSGRRHCLLAVCLCFCLNFAAANVNEDDNKEAAAALCGILEL
                     GAGRAKITPSTGLQTATYDEIQDLNMSLADAAWRSLFRDPSNQDNFRGFPTEEFGEST
                     DWKDKWEEWKNSATRLKEEAVLKQKLKAAGLEGASPSAMRHAQEIIAEIAEAAAHLRR
                     TTEEATKGKIIDQQAVQQKIDEAIYGEKIADEKAFGRAKVFNNAGGSRQANCEGNIGE
                     NKASTTLATLICLCAADNNGQTGSEHKACSGQTAVTQQWSGAAAPEQTTVTEMIQLCD
                     TKDSHQITATALQTRLEAVARQLRIINGAAYYGKFVAGNCNGEQGGGLCVKYTDINNN
                     AGKGFNSIPWVDKLRQLREQLEEHERAATKIEQTNTALNCAAAATKAIGRRVQLREAA
                     GSNAAEPVVTQKSAKSEGKQKECNAAGDDPKKCKDLEGKGCTYDEAKPKGQKCTLSED
                     AKKEVQQTAEKAAGTVSKNESKCGDKKTEGDCKDGCKWENNACKDSSILLNNQFALSV
                     VSAAFAALLF"
BASE COUNT    10433 a   6463 c   8050 g   9941 t
ORIGIN


  Query Match           55.0%;  Score 22;  DB 56;  Length 34887;
  Best Local Similarity  73.7%;  Pred. No. 60;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      3 cgagccaccccataaccctcaatactccagggattggg 40
          | |||||||| | |||| ||||| |||||  | |  ||
Db   8362 CAAGCCACCCAAGAACCATCAATTCTCCACTGTTATGG 8325


RESULT  14
HS821D11/c
LOCUS       HS821D11    76727 bp    DNA             PRI       12-DEC-1999
DEFINITION  Human DNA sequence from clone RP5-821D11 on chromosome 22q12.3-13.1
            Contains three partial unknown genes, one downstream of a predicted
            CpG island, and the first coding exon of the SREBF2 gene for Sterol
            Regulatory Element Binding Transcription Factor 2 downstream of a
            predicted CpG island, ESTs, STSs, GSSs and genomic marker D22S1157,
            complete sequence.
ACCESSION   AL021453
VERSION     AL021453.1  GI:3413288
KEYWORDS    HTG; D22S1157; SREBF2; Sterol Regulatory Element Binding
            Transcription Factor 2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 76727)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 1998 this sequence version replaced gi:3355590.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP5-821D11 is
            from the library RPCI-5 constructed at the Roswell Park Cancer
            Institute by the group of Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP5-821D11 It may be shorter because we only sequence overlapping
            sections once, or longer because we arrange for a small overlap
            between neighbouring submissions.
            The true left end of clone RP5-821D11 is at 1 in this sequence. The
            true left end of clone CTA-250D10 is at 76628 in this sequence. The
            true right end of clone CTA-109G6 is at 42082 in this sequence. The
            start of this sequence overlaps with sequence Z99716 The end of
            this sequence overlaps with sequence Z83840.
FEATURES              Location/Qualifiers
     source           1. .76727
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="22"
                      /map="q12.3-13.1"
                      /clone="RP5-821D11"
                      /clone_lib="RPCI-5"
     repeat_region    350. .434
                      /note="L2 repeat: matches 2490. .2537 of consensus"
     repeat_region    435. .743
                      /note="AluJb repeat: matches 8. .312 of consensus"
     variation        446. .448
                      /note="clone CTA-109G6
                      tct in this entry
                      substitution"
                      /replace="ttt"
     repeat_region    744. .840
                      /note="L2 repeat: matches 2381. .2490 of consensus"
     mRNA             join(1230. .1711,13695. .13736)
                      /gene="dJ821D11.1"
                      /note="match: ESTs: Em:AA419437"
                      /evidence=not_experimental
                      /product="dJ821D11.1 (PUTATIVE protein)"
     gene             1230. .13736
                      /gene="dJ821D11.1"
     variation        1314. .1316
                      /gene="dJ821D11.1"
                      /note="clone CTA-109G6
                      tgt in this entry
                      substitution"
```

```
                /replace="tct"
CDS             join(1601. .1711,13695. .>13736)
                /gene="dJ821D11.1"
                /note="other possible startcodon at 1583
                this gene and dJ821D11.2 could be part of one gene
                match: proteins: Tr:O95505"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ821D11.1 (PUTATIVE protein)"
                /protein_id="CAA16279.1"
                /db_xref="GI:4200330"
                /db_xref="SPTREMBL:O95505"
                /translation="MLALTLAKADSPRTALLCSAWLLTASFSAQQHKGSLQKDPLLSQ
                ACVGCLE"
variation       1760. .1762
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                gag in this entry
                substitution"
                /replace="ggg"
repeat_region   1773. .1923
                /note="MIR repeat: matches 47. .201 of consensus"
variation       1901. .1903
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                tca in this entry
                substitution"
                /replace="tta"
repeat_region   2010. .2209
                /note="L2 repeat: matches 2250. .2453 of consensus"
variation       2302. .2304
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                gga in this entry
                substitution"
                /replace="gaa"
repeat_region   3042. .3152
                /note="L1MC1 repeat: matches 6231. .6332 of consensus"
repeat_region   3162. .3457
                /note="AluSp repeat: matches 1. .297 of consensus"
repeat_region   3458. .3551
                /note="MIR repeat: matches 74. .148 of consensus"
repeat_region   3552. .3848
                /note="AluJb repeat: matches 1. .299 of consensus"
repeat_region   3849. .3887
                /note="MIR repeat: matches 32. .74 of consensus"
variation       4753. .4754
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                tt in this entry
                deletion"
                /replace="ttgagt"
repeat_region   4772. .4790
                /note="L1M4 repeat: matches 3167. .3185 of consensus"
repeat_region   4791. .5099
                /note="AluSq repeat: matches 2. .312 of consensus"
variation       4791. .4792
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                tt in this entry
                deletion"
                /replace="ttttatttat"
repeat_region   5100. .5116
                /note="L1M4 repeat: matches 3151. .3167 of consensus"
variation       5107. .5112
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                ttttat in this entry
                insertion"
                /replace="tt"
repeat_region   5117. .5427
                /note="AluYb8 repeat: matches 1. .318 of consensus"
variation       5359. .5361
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                att in this entry
                substitution"
                /replace="act"
repeat_region   5430. .5729
                /note="AluSp repeat: matches 1. .300 of consensus"
repeat_region   5730. .5764
                /note="L1M4 repeat: matches 3116. .3151 of consensus"
variation       5761. .5762
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                at in this entry
                deletion"
                /replace="att"
repeat_region   5765. .6075
                /note="AluY repeat: matches 1. .311 of consensus"
variation       5772. .5773
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                tt in this entry
                deletion"
                /replace="tat"
variation       5778. .5779
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                tt in this entry
                deletion"
                /replace="ttat"
variation       5780. .5782
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                tat in this entry
                substitution"
                /replace="ttt"
variation       5799. .5801
                /gene="dJ821D11.1"
                /note="clone CTA-109G6
                cgg in this entry
                substitution"

  Query Match          55.0%;  Score 22;  DB 77;  Length 76727;
  Best Local Similarity   73.7%;  Pred. No. 59;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 ctcgagccacccccataaccctcaatactccagggattg 38
          || ||| ||||| | | |||||   || |||||| |||
Db  34580 CTTGAGACACCCAACACCCCTCCCGACACCAGGGTTTG 34543


RESULT  15
HSAJ9611/c
LOCUS       HSAJ9611   110387 bp    DNA             HTG       07-OCT-2000
DEFINITION  Homo sapiens chromosome 19 clone PAC RPCI-1 155I10, *** SEQUENCING
            IN PROGRESS ***, 3 unordered pieces.
ACCESSION   AJ009611
VERSION     AJ009611.5  GI:9211527
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 110387)
  AUTHORS   Radelof,U., Hennig,S., Ramser,J., Francis,F., Steffens,C.,
            Klein,M., Seranski,P., Poustka,A., Reinhardt,R. and Lehrach,H.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 110387)
  AUTHORS   MPIMG.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-1998) MPIMG, Abt.Lehrach, Max Planck Institut
            fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany
COMMENT     On Jul 15, 2000 this sequence version replaced gi:8248730.
            contig 01          1. .40399
```

```
           contig 02      40500. .61762
           contig 03      61863. .110387
           Clone received from the Resource Centre of the Human Genome Project
           at the Max-Planck-Institute for Molecular Genetics.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 3 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *        1    40399: contig of 40399 bp in length
           *    40400 40499: gap of      100 bp
           *    40500    61762: contig of 21263 bp in length
           *    61763 61862: gap of      100 bp
           *    61863   110387: contig of 48525 bp in length.
FEATURES             Location/Qualifiers
     source          1. .110387
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="PAC RPCI-1 155I10"
                     /clone_lib="RPCI1,3-5 Human PAC library, originating
                     insitute: Roswell Park Cancer Institute, creator: Pieter
                     de Jong, P. Ioannou"
     misc_feature    1
                     /note="T7_end:PAC RPCI-1 155I10"
     misc_feature    110387
                     /note="SP6_end:PAC RPCI-1 155I10"
BASE COUNT    26958 a  27876 c  28291 g  27062 t    200 others
ORIGIN


  Query Match             55.0%;  Score 22;  DB 84;  Length 110387;
  Best Local Similarity   73.7%;  Pred. No. 59;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 tcgagccaccccataaccctcaatactccagggattgg 39
           | ||| || ||||| |||| ||| ||||||| |   ||
Db   15347 TGGAGTCATCCCATGACCCCCAAGACTCCAGTGCCAGG 15310


Search completed: January 18, 2001, 04:21:48
Job time: 10474 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 11:04:24 ; Search time 262.85 Seconds
(without alignments)
57.168 Million cell updates/sec

Title: PCT-US00-32259-12
Perfect score: 40
Sequence: 1 ctcgagccaccccataaccctcaatactccagggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*
1: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	23	57.5	2634	19	V59067	Bax inhibitor BI-1
c	2	20.8	52.0	837	20	X37525	Human secreted pro
c	3	20	50.0	3364	19	V07266	Muscle creatine ki
	4	19.8	49.5	2832	15	Q62186	Cyanobacteria repl
c	5	19.8	49.5	4809	15	Q62185	pAQ1 plasmid fragm
c	6	19.4	48.5	674	20	X37371	Human secreted pro
c	7	19.4	48.5	1554	20	X16675	Xenopus WA545 prot
	8	19.2	48.0	300	20	Z14673	Human gene express
c	9	19.2	48.0	972	19	V37941	Rat arginase encod
	10	19.2	48.0	1742	21	Z49765	Human DAP-kinase-r
	11	19.2	48.0	1864	21	Z99730	cDNA encoding huma
c	12	19.2	48.0	2302	20	X86274	DNA encoding a hum
	13	19	47.5	2401	14	Q50146	Restriction enzyme
	14	18.8	47.0	1183	19	X14116	H. pylori GHPO 106
c	15	18.8	47.0	2191	21	Z92219	Human thioredoxin
	16	18.8	47.0	32207	20	V73805	KSHV LUR DNA (nucl
	17	18.8	47.0	137507	19	V19941	KSHV long unique c
c	18	18.6	46.5	199	16	T19259	Human gene signatu
c	19	18.6	46.5	798	19	V30949	Human spermidine/s
c	20	18.6	46.5	2532	20	Z52953	Human prostate tum
c	21	18.6	46.5	21170	20	X20535	Polynucleotide seq
	22	18.4	46.0	453	20	V86886	EST clone BI3. Ho
	23	18.4	46.0	722	19	V28650	Ripening banana pu
c	24	18.4	46.0	1045	18	T59977	5' untranslated re
c	25	18.4	46.0	1045	20	X02016	D. melanogaster ti
	26	18.4	46.0	1194	21	Z91868	Streptococcus pneu
c	27	18.4	46.0	2082	15	Q55789	Sequence comprisin
c	28	18.4	46.0	2162	19	V37380	Streptococcus pneu
c	29	18.4	46.0	2382	12	Q10238	Encodes membrane e
	30	18.4	46.0	2425	13	Q30133	ERR receptor gene.
	31	18.4	46.0	2425	20	V59845	Murine ecotropic r
c	32	18.4	46.0	3954	18	T59975	Full length tipE p
c	33	18.4	46.0	3954	20	X02015	D. melanogaster ti
	34	18.4	46.0	21338	19	V52153	Streptococcus pneu
c	35	18.2	45.5	259	17	T07319	Partial formyl-CoA
c	36	18.2	45.5	259	20	V64283	O. formigenes form
c	37	18.2	45.5	870	19	Z96201	S. pneumoniae deri
	38	18.2	45.5	1233	20	Z08861	Human DNAX soluble
	39	18.2	45.5	1281	20	X89654	cDNA encoding a hu
	40	18.2	45.5	1305	20	Z06348	Nucleotide sequenc
	41	18.2	45.5	1347	21	Z50347	Human orphan cytok
	42	18.2	45.5	1391	19	V27144	Nucleotide sequenc
	43	18.2	45.5	1498	21	Z34614	Human receptor mol
	44	18.2	45.5	1579	19	V41689	Nucleotide sequenc
	45	18.2	45.5	1579	20	X90753	Human U4 haematopo

ALIGNMENTS

RESULT 1
V59067/c
ID V59067 standard; cDNA; 2634 BP.
XX
AC V59067;
XX
DT 02-FEB-1999 (first entry)
XX
DE Bax inhibitor BI-1 cDNA.
XX
KW Bax inhibitor; BI-1; human; apoptosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 73..786
FT /*tag= a
XX
PN WO9840397-A1.
XX
PD 17-SEP-1998.
XX
PF 13-MAR-1998; 98WO-US05015.
XX
PR 14-MAR-1997; 97US-0818514.
XX
PA (BURN-) BURNHAM INST.
XX
PI Reed JC, Xu Q;
XX
DR WPI; 1998-531519/45.
DR P-PSDB; W73136.
XX
PT Bax inhibitor proteins, BI-1 and BI-2 - useful e.g. to modulate
PT cellular apoptotic activity or identify agents altering BI-1 or BI-2

```
PT   binding which can modulate apoptotic activity
XX
PS   Claim 2; Page 61-63; 80pp; English.
XX
CC   This cDNA clone codes for an inhibitor protein, termed BI-1 (see
CC   W73136), of the pro-apoptotic protein Bax.  Nucleic acids encoding
CC   BI-1 and BI-2 (see V59068) were identified by suppression of
CC   Bax-induced death of yeast cells transformed to express human Bax.
CC   A human HepG2 cDNA library was used for library screening.  The
CC   invention provides vectors, optionally expression or viral vectors,
CC   containing BI nucleic acids, and host cells containing these
CC   vectors.  The nucleic acids encoding BI-1/BI-2 can be used to
CC   increase expression of these proteins in cells, or antisense
CC   molecules prepared from them used to decrease expression.  In
CC   these ways, cellular apoptotic activity may be modulated (claimed).
CC   The nucleic acids and complementary sequences are also useful as
CC   probes to detect BI-encoding nucleic acid molecules in samples.
XX
SQ   Sequence 2634 BP; 632 A; 626 C; 597 G; 779 T; 0 other;


  Query Match             57.5%;  Score 23;  DB 19;  Length 2634;
  Best Local Similarity   74.4%;  Pred. No. 0.92;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db   2056 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 2018


RESULT   2
X37525/c
ID   X37525 standard; cDNA; 837 BP.
XX
AC   X37525;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Human secreted protein cDNA fragment containing gene 75.
XX
KW   Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW   gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW   immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW   renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW   cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW   pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW   arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918208-A1.
XX
PD   15-APR-1999.
XX
PF   01-OCT-1998;   98WO-US20775.
XX
PR   02-OCT-1997;   97US-0060884.
PR   02-OCT-1997;   97US-0060833.
PR   02-OCT-1997;   97US-0060836.
PR   02-OCT-1997;   97US-0060837.
PR   02-OCT-1997;   97US-0060838.
PR   02-OCT-1997;   97US-0060839.
PR   02-OCT-1997;   97US-0060843.
PR   02-OCT-1997;   97US-0060862.
PR   02-OCT-1997;   97US-0060866.
PR   02-OCT-1997;   97US-0060874.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Carter KC,  Duan DR,  Endress GA,  Feng P,  Ferrie AM;
PI   Florence KA,  Greene JM,  Janat F,  Lafleur DW,  Ni J;
PI   Rosen CA,  Ruben SM,  Shi Y,  Young P,  Yu G;
XX
DR   WPI; 1999-264022/22.
DR   P-PSDB; Y07926.
XX
PT   New isolated human genes and the secreted polypeptides they encode
XX
PS   Claim 1a; Page 264; 368pp; English.
XX
CC   This invention describes novel isolated human genes and the secreted
CC   proteins they encode. The products of the invention are useful for
CC   preventing, treating or ameliorating medical conditions, e.g. by protein
CC   or gene therapy. Also pathological conditions can be diagnosed by
CC   determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 101 polynucleotides, based on
CC   which tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, tumours,
CC   neurodegenerative disorders, developmental abnormalities and fetal
CC   deficiencies, blood disorders, leukemias, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC   allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC   disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC   transplant rejection, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC   infections and AIDS. The human secreted proteins of the invention are
CC   represented in Y07852-Y07993 and the encoding nucleic acids are
CC   represented in X37451-X37552.
XX
SQ   Sequence 837 BP; 171 A; 249 C; 187 G; 228 T; 2 other;


  Query Match             52.0%;  Score 20.8;  DB 20;  Length 837;
  Best Local Similarity   70.0%;  Pred. No. 6.5;
  Matches   28;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 ctcgagccacccccataaccctcaatactccagggattggg 40
            |||  ||  ||||| ||| ||| | || |  | |||||||
Db    250 CTCAGGCAGCCCCAGAACACTCCAGACGCTGGAGATTGGG 211


RESULT   3
V07266/c
ID   V07266 standard; DNA; 3364 BP.
XX
AC   V07266;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Muscle creatine kinase enhancer/promoter sequence.
XX
KW   Encapsidated adenovirus minichromosome; EAM; gene therapy;
KW   vector; muscle creatine kinase; promoter; mouse; dystrophin;
KW   muscular dystrophy; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   enhancer        2102..2307
FT                   /*tag= a
FT   prim_transcript 3358..3364
FT                   /*tag= b
FT                   /note= "5' end of primary transcript"
XX
PN   WO9817783-A1.
XX
PD   30-APR-1998.
XX
PF   23-OCT-1997;   97WO-US19541.
XX
PR   23-OCT-1996;   96US-0735609.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
```

```
PI   Amalfitano A,  Chamberlain JS,  Hartigan-O'Connor DJ;
PI   Hauser MA,  Kumar-Singhr;
XX
DR   WPI; 1998-261485/23.
XX
PT   New adenoviral recombinant plasmid(s) - comprise sequences provided
PT   for expression of large foreign DNA fragments, used for, e.g. gene
PT   therapy of genetic disease(s)
XX
PS   Example 6a; Page 105-107; 139pp; English.
XX
CC   This nucleotide sequence comprises the promoter/enhancer region
CC   of the murine muscle creatine kinase (MCK) gene.  The invention
CC   provides encapsidated adenovirus minichromosomes (EAMs) containing
CC   a full-length murine dystrophin cDNA.  The EAM consists of an
CC   infectious encapsidated linear genome containing adenovirus
CC   origins of replication, packaging signal elements, a
CC   beta-galactosidase reporter gene cassette and a full-length
CC   (14 kb) dystrophin cDNA regulated by the MCK enhancer/promoter.
CC   EAMs are generated by cotransfecting 293 cells with supercoiled
CC   plasmid DNA (pAd5 beta-dys) containing an embedded inverted
CC   origin of replication (and the remaining above elements) with
CC   linear DNA from E1-deleted virions expressing human placental
CC   alkaline phosphatase (hpAP).  All proteins necessary for the
CC   generation of EAMs are provided in trans from the hpAP virions.
CC   The EAMs are useful for gene transfer to a variety of cell
CC   types in vitro and in vivo.  The invention provides improved
CC   adenovirus vectors (see also V07261) and packaging cell lines
CC   useful for a wide variety of gene therapy applications.
XX
SQ   Sequence 3364 BP; 860 A; 900 C; 810 G; 794 T; 0 other;


  Query Match             50.0%;  Score 20;  DB 19;  Length 3364;
  Best Local Similarity   72.2%;  Pred. No. 18;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       5 agccaccccataaccctcaatactccagggattggg 40
           ||||  ||| | |||||  || | |||||||  |||
Db    2326 AGCCTGCCCCTCACCCTGGATCCACCAGGGACAGGG 2291


RESULT   4
Q62186
ID   Q62186 standard; DNA; 2832 BP.
XX
AC   Q62186;
XX
DT   22-NOV-1994  (first entry)
XX
DE   Cyanobacteria replication protein coding sequence.
XX
KW   Plasmid; pAQ1; Cyanobacteria; PCC7002; vector; expression;
KW   carbon dioxide; air; ss.
XX
OS   Agmenellum quadruplicatum PR-6.
XX
PN   JP06078777-A.
XX
PD   22-MAR-1994.
XX
PF   02-SEP-1992;   92JP-0274768.
XX
PR   02-SEP-1992;   92JP-0274768.
XX
PA   (KANT ) KANSAI DENRYOKU KK.
XX
DR   WPI; 1994-131278/16.
DR   P-PSDB; R51499.
XX
PT   New vector from cyanobacteria. - used for foreign gene expression
PT   in cyanobacteria
XX
PS   Disclosure; Page 15-18; 23pp; Japanese.
XX
CC   This sequence represents a fragment of the plasmid pAQ1 and encodes
CC   a protein which has replication activity in Cyanobacteria.  This
CC   sequence was used in the production of a vector for the expression
CC   of heterologous proteins in Cyanobacteria.  This allows effective use
CC   of carbon dioxide in the air.
XX
SQ   Sequence 2832 BP; 760 A; 715 C; 724 G; 633 T; 0 other;


  Query Match             49.5%;  Score 19.8;  DB 15;  Length 2832;
  Best Local Similarity   69.2%;  Pred. No. 21;
  Matches   27;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       2 tcgagccaccccataaccctcaatactccagggattggg 40
           |||||| |||  | | |  ||||  || | |||| ||||
Db     305 tcgagctaccagaaaccgatcaaggcttctgggaatggg 343


RESULT   5
Q62185/c
ID   Q62185 standard; DNA; 4809 BP.
XX
AC   Q62185;
XX
DT   22-NOV-1994  (first entry)
XX
DE   pAQ1 plasmid fragment.
XX
KW   Plasmid; pAQ1; Cyanobacteria; PCC7002; vector; expression;
KW   carbon dioxide; air; ss.
XX
OS   Agmenellum quadruplicatum PR-6.
XX
PN   JP06078777-A.
XX
PD   22-MAR-1994.
XX
PF   02-SEP-1992;   92JP-0274768.
XX
PR   02-SEP-1992;   92JP-0274768.
XX
PA   (KANT ) KANSAI DENRYOKU KK.
XX
DR   WPI; 1994-131278/16.
XX
PT   New vector from cyanobacteria. - used for foreign gene expression
PT   in cyanobacteria
XX
PS   Disclosure; Page 9-14; 23pp; Japanese.
XX
CC   This sequence represents a fragment of the plasmid pAQ1.  This
CC   plasmid was prepared from Cyanobacteria PCC7002.  This sequence was
CC   used in the production of a vector for the expression of heterologous
CC   proteins in Cyanobacteria.  This allows effective use of carbon
CC   dioxide in the air.
XX
SQ   Sequence 4809 BP; 1136 A; 1193 C; 1165 G; 1315 T; 0 other;


  Query Match             49.5%;  Score 19.8;  DB 15;  Length 4809;
  Best Local Similarity   69.2%;  Pred. No. 24;
  Matches   27;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       2 tcgagccaccccataaccctcaatactccagggattggg 40
           |||||| |||  | | |  ||||  || | |||| ||||
Db    2348 TCGAGCTACCAGAAACCGATCAAGGCTTCTGGGAATGGG 2310


RESULT   6
X37371/c
ID   X37371 standard; cDNA; 674 BP.
```

```
XX
AC   X37371;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Human secreted protein cDNA fragment containing gene 3.
XX
KW   Human; secreted protein; prevention; treatment; protein therapy;
KW   gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; foetal deficiency; blood disorder; lymphoma;
KW   leukemia; immune system disorder; autoimmune disease; hepatic disease;
KW   renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;
KW   Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;
KW   arthritis; psoriasis; digestive; endocrine; infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO9909155-A1.
XX
PD   25-FEB-1999.
XX
PF   18-AUG-1998;   98WO-US17044.
XX
PR   16-JUN-1998;   98US-0092956.
PR   15-JUL-1998;   98US-0092956.
PR   19-AUG-1997;   97US-0056368.
PR   19-AUG-1997;   97US-0056369.
PR   19-AUG-1997;   97US-0056535.
PR   19-AUG-1997;   97US-0056555.
PR   19-AUG-1997;   97US-0056556.
PR   19-AUG-1997;   97US-0056628.
PR   19-AUG-1997;   97US-0056629.
PR   19-AUG-1997;   97US-0056726.
PR   19-AUG-1997;   97US-0056728.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Duan R,  Ebner R,  Endress GA,  Feng P;
PI   Florence C,  Florence KA,  Komatsoulis GA,  Lafleur DW;
PI   Moore PA,  Olsen HS,  Rosen CA,  Ruben SM,  Shi Y,  Soppet DR;
PI   Young PE;
XX
DR   WPI; 1999-190160/16.
DR   P-PSDB; Y07746.
XX
PT   New isolated human genes and the secreted polypeptides they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1a; Page 192; 280pp; English.
XX
CC   This invention describes novel isolated human secreted proteins and
CC   their encoding nucleic acid sequences. The products of the invention
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also pathological conditions can be
CC   diagnosed by determining the presence or amount of expression of
CC   the new polypeptides in a sample or by determining the presence or
CC   absence of mutations in the new polynucleotides. Specific uses are
CC   described for each of the 70 polynucleotides, based on which tissues they
CC   are most highly expressed in, and include developing products for the
CC   diagnosis or treatment of cancer, tumours, neurodegenerative
CC   disorders, developmental abnormalities and foetal deficiencies, blood
CC   disorders, leukemias, diseases of the immune system, autoimmune diseases,
CC   hepatic and renal disease, lymphomas, inflammation, allergies, asthma,
CC   sepsis, diabetes, Alzheimer's and cognitive disorders, schizophrenia,
CC   osteoporosis, arthritis, psoriasis, digestive/endocrine disorders,
CC   infections and AIDS. The human secreted proteins of the invention are
CC   represented in Y07744-Y07850 and the encoding nucleic acids are
CC   represented in X37369-X37441.
XX
SQ   Sequence 674 BP; 238 A; 112 C; 175 G; 145 T; 4 other;

  Query Match          48.5%;  Score 19.4;  DB 20;  Length 674;
  Best Local Similarity   70.3%;  Pred. No. 24;
  Matches   26;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 gagccaccccataaccctcaatactccagggattggg 40
          ||||  ||| .| ||| |||||||    || ||||||
Db    645 GAGCTCCCCACTCACCTTCAATACAAATGGCATTGGG 609

RESULT   7
X16675/c
ID   X16675 standard; DNA; 1554 BP.
XX
AC   X16675;
XX
DT   29-APR-1999  (first entry)
XX
DE   Xenopus WA545 protein encoding DNA.
XX
KW   Xenopus; WA545 protein; TGF-beta; transforming growth factor beta;
KW   proliferation; differentiation; mesodermal tissue; neural; muscle;
KW   bone; cartilage; connective tissue; wound healing; gene therapy; ss.
XX
OS   Xenopus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             55..1119
FT                   /*tag= a
FT                   /product= "WA545 protein"
FT   sig_peptide     55..774
FT                   /*tag= b
FT   mat_peptide     775..1116
FT                   /*tag= c
XX
PN   WO9902678-A1.
XX
PD   21-JAN-1999.
XX
PF   24-APR-1998;   98WO-US08334.
XX
PR   10-JUL-1997;   97US-0890918.
XX
PA   (GEMY ) GENETICS INST INC.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Lavallie ER,  Racie LA,  Sive H,  Sun B;
XX
DR   WPI; 1999-120879/10.
DR   P-PSDB; W94656.
XX
PT   New isolated Xenopus WA545 DNA - which is a member of the
PT   transforming growth factor-beta superfamily, used to induce tissue
PT   formation and in wound healing
XX
PS   Claim 1; Page 51-52; 73pp; English.
XX
CC   The present sequence encodes the Xenopus WA545 protein. WA545 proteins
CC   have the ability to induce, enhance and/or inhibit the formation,
CC   growth, proliferation, differentiation, maintenance of mesodermal
CC   tissue, including neurons and/or related neural cells and tissues such
CC   as brain cells, Schwann cells, glial cells and astrocytes, as well as
CC   muscle cells and tissues. They can be used for treating bone, cartilage,
CC   muscle, nerve, epidermis and/or other connective tissue defects, as well
CC   as periodontal disease and healing of various epidermis, nerve,
CC   including spinal cord, muscle, including cardiac, striated or smooth
CC   muscle, and other tissues and wounds, and other organs such as liver,
CC   pancreas, spleen brain, lung, cardiac and kidney tissue. They can also
CC   be used to treat or prevent such conditions as osteoarthritis,
CC   osteoporosis, and other abnormalities of bone, cartilage, muscle, nerve,
CC   epidermis or other connective tissue, organs such as liver, pancreas,
CC   spleen, lung, cardiac and kidney and other tissues. They can also be
CC   used for wound healing, reduction of fibrosis and reduction of scar
CC   tissue formation. They can also be used to induce bone and/or cartilage
```

CC or other connective tissue formation, and in wound healing and tissue
CC repair and for augmenting the activity of bone morphogenetic proteins.
CC They can also be used as a dietary or nutrient supplement. The products
CC can also be used for treating cell populations, such as embryonic cells
CC or stem cell populations to enhance or enrich the growth and/or
CC differentiation of the cells and for implantation and for gene therapy
CC applications.
XX
SQ Sequence 1554 BP; 503 A; 340 C; 309 G; 402 T; 0 other;

Query Match 48.5%; Score 19.4; DB 20; Length 1554;
Best Local Similarity 70.3%; Pred. No. 28;
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 tcgagccaccccataaccctcaatactccagggattg 38
|| || || || |||| || || | |||||||||
Db 279 TCCAGGTACTCCGAAACCTTCCATCATGCAGGGATTG 243

RESULT 8
Z14673
ID Z14673 standard; cDNA; 300 BP.
XX
AC Z14673;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:2142.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 1144; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in Z12532 to Z17779. Also described is a
CC method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in Z12532 to Z17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 300 BP; 38 A; 122 C; 91 G; 49 T; 0 other;

Query Match 48.0%; Score 19.2; DB 20; Length 300;
Best Local Similarity 67.5%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 ctcgagccaccccataaccctcaatactccagggattggg 40
| || ||| ||| ||||| |||| | || | |||||
Db 222 cacgcgccgacccccaaccccccaattccccggcacttggg 261

RESULT 9
V37941/c
ID V37941 standard; DNA; 972 BP.
XX
AC V37941;
XX
DT 07-OCT-1998 (first entry)
XX
DE Rat arginase encoding DNA.
XX
KW Rat; tissue fibrosis inhibitor; arginase; blood; lung; kidney;
KW liver; skin; pancreas; ss.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..972
FT /*tag= a
FT /product= "arginase"
XX
PN WO9824473-A1.
XX
PD 11-JUN-1998.
XX
PF 03-DEC-1997; 97WO-JP04423.
XX
PR 03-DEC-1996; 96JP-0323105.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Lee J, Mitsui H, Sato M, Sugimoto S, Yano K;
XX
DR WPI; 1998-333049/29.
DR P-PSDB; W62042.
XX
PT Tissue fibrosis inhibitor containing an arginase or modified
PT arginase - and method for production of a recombinant arginase for
PT this use
XX
PS Claim 13; Page 64-66; 92pp; Japanese.
XX
CC The present sequence encodes rat arginase. The present invention
CC describes a tissue fibrosis inhibitor composition which contains as
CC an active component an enzyme which can decompose arginine, the enzyme
CC comprises human or rat arginase (e.g. derived from liver tissue) or
CC arginine deimidase or their modified derivatives (such as polyethylene
CC glycol-modified enzyme). The invention can be used in the prevention
CC and treatment of fibrosis in tissues such as lung, kidney, liver, skin

```
CC   or pancreas. The arginase has a high stability and long-lasting
CC   activity in the blood and is active in a wide range of tissues.
XX
SQ   Sequence 972 BP; 276 A; 225 C; 268 G; 203 T; 0 other;


  Query Match             48.0%;  Score 19.2;  DB 19;  Length 972;
  Best Local Similarity   67.5%;  Pred. No. 32;
  Matches   27;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
          || |  |||  |||| | | |||| ||||  | || ||||
Db    715 CTGGGTCCAGTCCATCAACATCAAAACTCAGGTGAATGGG 676


RESULT  10
Z49765
ID   Z49765 standard; DNA; 1742 BP.
XX
AC   Z49765;
XX
DT   18-APR-2000  (first entry)
XX
DE   Human DAP-kinase-related protein 1 (DRP-1) encoding DNA.
XX
KW   DAP-kinase-related protein 1; DRP-1; Death-Associated Protein;
KW   calmodulin-dependent serine/threonine kinase; apoptosis; dimerisation;
KW   cytostatic; antipsoriatic; immunosuppressive; metastasis; tumour; human;
KW   treatment; cancer; psoriasis; autoimmune disease; screening; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             62..1144
FT                   /*tag= a
FT                   /product= "DAP-kinase-related protein 1"
FT   polyA_signal    1720..1725
FT                   /*tag= b
XX
PN   WO9966030-A1.
XX
PD   23-DEC-1999.
XX
PF   15-JUN-1999;   99WO-US13411.
XX
PR   15-JUN-1998;   98US-0089294.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (MCIN/) MCINNIS P A.
XX
PI   Kimchi A;
XX
DR   WPI; 2000-147148/13.
DR   P-PSDB; Y44674.
XX
PT   Calmodulin-dependent serine/threonine kinase capable of inducing
PT   apoptosis used in the treatment of e.g. cancer -
XX
PS   Claim 3; Fig 1; 67pp; English.
XX
CC   The present sequence encodes DAP (death-associated protein)-
CC   kinase-related protein 1 (DRP-1), which is a calmodulin-dependent
CC   serine/threonine kinase. DRP-1 is a cytoplasmic protein capable of
CC   inducing apoptosis by dimerisation. It shows significant homology to
CC   DAP kinase. It has cytostatic, antipsoriatic and immunosuppressive
CC   activity and can be used for inhibiting growth/metastasis of
CC   tumours and promoting death of tumour cells. It can also be used in
CC   the treatment of cancer, psoriasis and autoimmune diseases. Fragments
CC   of DRP-1 DNA are useful as probes for screening individuals with a
CC   predisposition to cancer.
XX
SQ   Sequence 1742 BP; 442 A; 437 C; 504 G; 359 T; 0 other;


  Query Match             48.0%;  Score 19.2;  DB 21;  Length 1742;
  Best Local Similarity   67.5%;  Pred. No. 35;
  Matches   27;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
          ||| |||   |||| | | || ||| |||||   | ||||
Db   1342 ctccagcattcccaaagctcttaattctccataaaatggg 1381


RESULT  11
Z99730
ID   Z99730 standard; cDNA; 1864 BP.
XX
AC   Z99730;
XX
DT   12-JUL-2000  (first entry)
XX
DE   cDNA encoding human cardiovascular system associated protein kinase-4.
XX
KW   Human; cardiovascular system associated protein kinase-4; CSAPK-4;
KW   signalling pathway; cell growth; cell differentiation; gene mapping;
KW   tissue typing; forensic identification; cardiovascular disease;
KW   congestive heart failure; transgenic animal; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             275..757
FT                   /*tag= a
FT                   /product= " cardiovascular system associated protein
FT                            kinase-4"
XX
PN   WO200014212-A1.
XX
PD   16-MAR-2000.
XX
PF   09-SEP-1999;   99WO-US20631.
XX
PR   09-SEP-1998;   98US-0099657.
PR   29-SEP-1998;   98US-0163115.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Acton S;
XX
DR   WPI; 2000-271053/23.
DR   P-PSDB; Y84323.
XX
PT   New nucleic acid encoding cardiovascular system associated protein
PT   kinase, used e.g. for diagnosis, treatment and prevention of
PT   cardiovascular disease -
XX
PS   Claim 1; Fig 4; 163pp; English.
XX
CC   The present sequence encodes a human cardiovascular system associated
CC   protein kinase-4 (CSAPK-4). CSAPK polypeptides are involved in signalling
CC   pathways associated with cell growth and differentiation. The CSAPK
CC   polypeptides and polynucleotides are used to screen for agents that
CC   specifically modulate CSAPK, which are potential therapeutic agents.
CC   They are also used for diagnosis, prognosis or monitoring of
CC   CSAPK-related diseases, gene mapping, tissue typing and forensic
CC   identification, and for treating or preventing disorders associated
CC   with aberrant CSAPK expression or activity, especially cardiovascular
CC   diseases such as congestive heart failure. They can also be used in
CC   pharmacogenomics. The CSAPK polynucleotide may also be used to generate
CC   transgenic animals.
XX
SQ   Sequence 1864 BP; 437 A; 504 C; 527 G; 396 T; 0 other;


  Query Match             48.0%;  Score 19.2;  DB 21;  Length 1864;
  Best Local Similarity   67.5%;  Pred. No. 36;
```

```
  Matches   27;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
          ||| |||   |||| | | || ||| |||||   | ||||
Db    955 ctccagcattcccaaagctcttaattctccataaaatggg 994


RESULT  12
X86274/c
ID   X86274 standard; DNA; 2302 BP.
XX
AC   X86274;
XX
DT   22-SEP-1999  (first entry)
XX
DE   DNA encoding a human PIG protein.
XX
KW   p53 transcription tag; p53 status; cancer; cytotoxicity; carcinogenicity;
KW   neoplastic; PIG; ss.
XX
OS   Homo sapiens.
XX
PN   WO9914356-A2.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US19300.
XX
PR   30-MAR-1998;   98US-0079817.
PR   17-SEP-1997;   97US-0059153.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Kinzler KW,  Polyak K,  Vogelstein B;
XX
DR   WPI; 1999-443793/37.
XX
PT   Use of p53 transcription tags to determine p53 status in, e.g.
PT   cancer diagnosis
XX
PS   Disclosure; Page 65-66; 73pp; English.
XX
CC   The specification describes the use of p53 transcription tags for
CC   developing products to determine p53 status, to diagnose cancer
CC   and to evaluate cytotoxicity or carcinogenicity of a test agent.
CC   A method for diagnosing cancer or determining p53 status in a sample
CC   suspected for being neoplastic comprises comparing the level of
CC   transcription of an RNA transcript in a first sample (s1) of a first
CC   tissue (t1) to the level of transcription of the transcript in a second
CC   sample (s2) of a second tissue (s2), where s1 is suspected of being
CC   neoplastic and s2 is a normal human tissue (of the same type) and the
CC   transcript is identified by a tag; and categorizing s1 as neoplastic
CC   or as having a mutant p53 when transcription is found to be the same
CC   or lower in the first, than in s2. The methods and products can be used
CC   to determine p53 status, to diagnose cancer and to evaluate cytotoxicity
CC   or carcinogenicity of a test agent. X86265-75 encode human PIG proteins.
XX
SQ   Sequence 2302 BP; 497 A; 752 C; 594 G; 459 T; 0 other;


  Query Match            48.0%;  Score 19.2;  DB 20;  Length 2302;
  Best Local Similarity  75.0%;  Pred. No. 37;
  Matches   24;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccag 32
          |||||   | ||||  |||||||  |||||||
Db     42 CTCGACTTAGCCCACTACCCTCATGACTCCAG 11


RESULT  13
Q50146
ID   Q50146 standard; DNA; 2401 BP.
XX
AC   Q50146;
XX
DT   03-JUN-1994  (first entry)
XX
DE   Restriction enzyme.
XX
KW   Restriction site; enzyme; SphI; BamHI; cleavage-site;
KW   phage; Escherichia coli; ss.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     240..1074
FT                   /*tag= b
FT   misc_feature    897
FT                   /*tag= c
FT                   /label= SphI_cleavage-site
FT   mat_peptide     1127..2008
FT                   /*tag= d
FT   misc_feature    1425
FT                   /*tag= e
FT                   /label= BamHI_cleavage-site
XX
PN   JP05244946-A.
XX
PD   24-SEP-1993.
XX
PF   03-MAR-1992;   92JP-0094990.
XX
PR   03-MAR-1992;   92JP-0094990.
XX
PA   (KAJI/) KAJI A.
XX
DR   WPI; 1993-389087/49.
DR   P-PSDB; R42305.
XX
PT   New restriction enzyme - suppresses melting by T-even phage of
PT   Escherichia coli
XX
PS   Claim 1; Page 7-10; 10pp; Japanese.
XX
CC   The enzyme can suppress the melting by T-even phage of E.coli, in a
CC   temp. dependent manner.  A large amt. of enzyme can be produced.
XX
SQ   Sequence 2401 BP; 681 A; 564 C; 543 G; 613 T; 0 other;


  Query Match            47.5%;  Score 19;  DB 14;  Length 2401;
  Best Local Similarity  71.4%;  Pred. No. 46;
  Matches   25;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      3 cgagccaccccataaccctcaatactccagggatt 37
          | || |||| || ||| | |||||| |  | ||||
Db    284 caaggcaccgcacaacacgcaatacacaggcgatt 318


RESULT  14
X14116
ID   X14116 standard; DNA; 1183 BP.
XX
AC   X14116;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 106 gene.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease; ss.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..1133
FT                   /*tag= a
XX
```

```
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   P-PSDB; W98397.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 1; Page 650-651; 2054pp; English.
XX
CC   This sequence represents a polynucleotide of the invention. It was
CC   isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC   The polypeptides can be used for preventing or treating Helicobacter
CC   infections, and gastroduodenal diseases associated with these
CC   infections, including acute, chronic, and atrophic gastritis, and peptic
CC   ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC   for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence 1183 BP; 413 A; 200 C; 242 G; 328 T; 0 other;


  Query Match             47.0%;  Score 18.8;  DB 19;  Length 1183;
  Best Local Similarity   68.4%;  Pred. No. 49;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattgg 39
          ||| |    | |||||||||| | ||||||   || |||
Db    355 tcgtgagcgctcataaccctaattactcctatgaatgg 392


RESULT  15
Z92219/c
ID   Z92219 standard; DNA; 2191 BP.
XX
AC   Z92219;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human thioredoxin reductase nucleotide sequence #1.
XX
KW   Thioredoxin reductase; redox inhibitor; cellular proliferation; cancer;
KW   apoptosis; reperfusion injury; hepatitis; amyotrophic lateral sclerosis;
KW   Alzheimer's disease; autoimmune disease; Sjogren's syndrome; lupus;
KW   rheumatoid arthritis; HIV; Hermansky-Pudlack syndrome; retinopathy;
KW   wound healing; Crohn's disease; ulcerative colitis; angiogenesis;
KW   adult respiratory distress syndrome; ARDS; infection; ds;
KW   inflammatory conditions, psoriasis; inflammatory bowel disease;
KW   hyperproliferative disease; mycosis fungoides.
XX
OS   Homo sapiens.
XX
PN   WO200006088-A2.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;   99WO-US17496.
XX
PR   31-JUL-1998;   98US-0127219.
XX
PA   (KIRK/) KIRKPATRICK D L.
PA   (POWI/) POWIS G.
XX
PI   Kirkpatrick DL,  Powis G;
XX
DR   WPI; 2000-195158/17.
XX
PT   New composition for preventing inhibition of apoptosis and for treating
PT   e.g. cancer or reperfusion injuries comprises inhibitor of cellular
PT   redox signalling -
XX
PS   Disclosure; Page 51-53; 55pp; English.
XX
CC   This sequence represents a thioredoxin reductase nucleotide sequence.
CC   Thioredoxin exerts specific redox control over a number of transcription
CC   factors to modulate their DNA binding and therefore regulates gene
CC   transcription. Thioredoxin reductase catalyses the reduction of
CC   thioredoxin. The thioredoxin redox system plays an important role in a
CC   number of diseased states. The invention relates to a composition
CC   comprising an inhibitor of cellular redox signalling and a carrier. The
CC   composition can be used for inhibiting cellular proliferation or
CC   preventing inhibition of apoptosis and is therefore useful for treating
CC   cancer, reperfusion injury following ischaemia, hepatitis, amyotrophic
CC   lateral sclerosis, neurodegenerative diseases, Alzheimer's disease,
CC   autoimmune diseases, Sjogren's syndrome, lupus, rheumatoid arthritis,
CC   HIV, Hermansky-Pudlack syndrome, retinal oxidative damage, retinopathy,
CC   skin hyperplasia, aging, ultraviolet damage, wound healing, Crohn's
CC   disease, ulcerative colitis, angiogenesis, uterine disorders, adult
CC   respiratory distress syndrome (ARDS), lung disorders, viral infections
CC   such as herpes virus, pox virus and adenovirus infections, inflammatory
CC   conditions, autoimmune diseases (such as systemic lupus erythematosus,
CC   rheumatoid arthritis, psoriasis, inflammatory bowel disease and
CC   autoimmune diabetes), immune mediated glomerular nephritis,
CC   hyperproliferative diseases (such as fibrosis) and mycosis fungoides.
XX
SQ   Sequence 2191 BP; 614 A; 411 C; 555 G; 611 T; 0 other;


  Query Match             47.0%;  Score 18.8;  DB 21;  Length 2191;
  Best Local Similarity   76.7%;  Pred. No. 55;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      5 agccaccccataaccctcaatactccaggg 34
          ||||| ||   |||||| | |||||||||
Db    878 AGCCAGCCAGAAACCCTGCACACTCCAGGG 849


Search completed: January 18, 2001, 11:04:27
Job time: 29436 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 10:59:51 ; Search time 153.62 Seconds
(without alignments)
41.963 Million cell updates/sec

Title: PCT-US00-32259-12
Perfect score: 40
Sequence: 1 ctcgagccaccccataaccctcaatactccagggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	23	57.5	2634	2	US-08-818-514-1	Sequence 1, Appli
2	23	57.5	2634	2	US-08-818-514-2	Sequence 2, Appli
c 3	23	57.5	2634	3	US-09-115-934A-1	Sequence 1, Appli
4	23	57.5	2634	3	US-09-115-934A-2	Sequence 2, Appli
c 5	20	50.0	3364	2	US-08-735-609-9	Sequence 9, Appli
c 6	20	50.0	3364	2	US-08-735-609-9	Sequence 9, Appli
c 7	20	50.0	3364	3	US-09-315-372-9	Sequence 9, Appli
c 8	20	50.0	3364	3	US-09-244-752-9	Sequence 9, Appli
c 9	20	50.0	3364	3	US-09-245-497-9	Sequence 9, Appli
10	19.2	48.0	1864	3	US-09-221-235-10	Sequence 10, Appl
11	19.2	48.0	1864	3	US-09-221-928-10	Sequence 10, Appl
12	19.2	48.0	1864	3	US-09-221-527-10	Sequence 10, Appl
13	19.2	48.0	1864	3	US-09-221-236-10	Sequence 10, Appl
14	19.2	48.0	1864	3	US-09-221-416-10	Sequence 10, Appl
15	18.8	47.0	32207	2	US-08-770-379-20	Sequence 20, Appl
c 16	18.6	46.5	798	2	US-08-742-009-2	Sequence 2, Appli
c 17	18.6	46.5	798	3	US-09-172-110-2	Sequence 2, Appli
c 18	18.4	46.0	1045	1	US-08-317-880-3	Sequence 3, Appli
c 19	18.4	46.0	1045	2	US-08-782-396-3	Sequence 3, Appli
20	18.4	46.0	2425	2	US-08-132-990A-3	Sequence 3, Appli
21	18.4	46.0	2425	4	PCT-US92-09382-3	Sequence 3, Appli
c 22	18.4	46.0	3954	1	US-08-317-880-1	Sequence 1, Appli
c 23	18.4	46.0	3954	2	US-08-782-396-1	Sequence 1, Appli
c 24	18.2	45.5	259	2	US-08-493-197-3	Sequence 3, Appli
c 25	18.2	45.5	259	4	PCT-US95-07844-3	Sequence 3, Appli
26	18.2	45.5	1305	3	US-09-012-072-3	Sequence 3, Appli
27	18.2	45.5	2584	3	US-08-758-662-8	Sequence 8, Appli
28	18.2	45.5	2784	1	US-08-474-140-10	Sequence 10, Appl
29	18.2	45.5	2784	1	US-08-477-630-10	Sequence 10, Appl
30	18.2	45.5	2784	1	US-08-472-293-10	Sequence 10, Appl
31	18.2	45.5	2784	1	US-08-474-545-10	Sequence 10, Appl
32	18.2	45.5	2784	1	US-08-478-341-10	Sequence 10, Appl
33	18.2	45.5	2784	3	US-08-996-733-10	Sequence 10, Appl
34	18.2	45.5	4464	1	US-08-474-140-8	Sequence 8, Appli
35	18.2	45.5	4464	1	US-08-474-140-9	Sequence 9, Appli
36	18.2	45.5	4464	1	US-08-477-630-8	Sequence 8, Appli
37	18.2	45.5	4464	1	US-08-477-630-9	Sequence 9, Appli
38	18.2	45.5	4464	1	US-08-472-293-8	Sequence 8, Appli
39	18.2	45.5	4464	1	US-08-472-293-9	Sequence 9, Appli
40	18.2	45.5	4464	1	US-08-474-545-8	Sequence 8, Appli
41	18.2	45.5	4464	1	US-08-474-545-9	Sequence 9, Appli
42	18.2	45.5	4464	1	US-08-478-341-8	Sequence 8, Appli
43	18.2	45.5	4464	1	US-08-478-341-9	Sequence 9, Appli
44	18.2	45.5	4464	3	US-08-996-733-8	Sequence 8, Appli
45	18.2	45.5	4464	3	US-08-996-733-9	Sequence 9, Appli

ALIGNMENTS

```
RESULT 1
US-08-818-514-1/c
; Sequence 1, Application US/08818514
; Patent No. 5837838
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Xu, Qunli
;    TITLE OF INVENTION:  BAX Inhibitor Proteins
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,514
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2446
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (519) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2634 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-818-514-1


  Query Match             57.5%;  Score 23;  DB 2;  Length 2634;
  Best Local Similarity   74.4%;  Pred. No. 0.58;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db   2056 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 2018


RESULT 2
```

```
US-08-818-514-2
; Sequence 2, Application US/08818514
; Patent No. 5837838
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Xu, Qunli
;    TITLE OF INVENTION:  BAX Inhibitor Proteins
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,514
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2446
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2634 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-818-514-2


  Query Match             57.5%;  Score 23;  DB 2;  Length 2634;
  Best Local Similarity   74.4%;  Pred. No. 0.58;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 tcgagccaccccataaccctcaatactccagggattggg 40
           || | |||| |   || || ||| ||| ||||||||||
Db     579 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 617


RESULT   3
US-09-115-934A-1/c
; Sequence 1, Application US/09115934A
; Patent No. 6130317
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Xu, Qunli
;    TITLE OF INVENTION:  BAX Inhibitor Proteins
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/115,934A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/818,514
;      FILING DATE:  14-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 3209
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2634 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-115-934A-1


  Query Match             57.5%;  Score 23;  DB 3;  Length 2634;
  Best Local Similarity   74.4%;  Pred. No. 0.58;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 tcgagccaccccataaccctcaatactccagggattggg 40
           || | |||| |   || || ||| ||| ||||||||||
Db    2056 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 2018


RESULT   4
US-09-115-934A-2
; Sequence 2, Application US/09115934A
; Patent No. 6130317
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Xu, Qunli
;    TITLE OF INVENTION:  BAX Inhibitor Proteins
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/115,934A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/818,514
;      FILING DATE:  14-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 3209
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2634 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-115-934A-2
```

```
  Query Match             57.5%;  Score 23;  DB 3;  Length 2634;
  Best Local Similarity   74.4%;  Pred. No. 0.58;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db    579 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 617


RESULT   5
US-08-735-609-9/c
; Sequence 9, Application US/08735609
; Patent No. 5955360
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar-Singh, Rajendra
;    APPLICANT:  Hartigan-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States Of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/735,609
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM-02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3364 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-735-609-9


  Query Match             50.0%;  Score 20;  DB 2;  Length 3364;
  Best Local Similarity   72.2%;  Pred. No. 11;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      5 agccaccccataaccctcaatactccagggattggg 40
          ||||  ||| | |||||  || | |||||||  |||
Db   2326 AGCCTGCCCCTCACCCTGGATCCACCAGGGACAGGG 2291


RESULT   6
US-08-735-609-9/c
; Sequence 9, Application US/08735609
; Patent No. 5994132
;  GENERAL INFORMATION:
;       APPLICANT: Chamberlain, Jeffrey S.
;                  Amalfitano, Andrea
;                  Hauser, Michael A.
;                  Kumar-Singh, Rajendra
;                  Hartigan-O'Connor, Dennis J.
;       TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
;       NUMBER OF SEQUENCES: 15
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll, LLP
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States Of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/735,609
;            FILING DATE: 23-Oct-1996
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: UM-02484
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 9:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 3364 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;            DESCRIPTION: /desc = "DNA"
;       SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-08-735-609-9


  Query Match             50.0%;  Score 20;  DB 2;  Length 3364;
  Best Local Similarity   72.2%;  Pred. No. 11;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      5 agccaccccataaccctcaatactccagggattggg 40
          ||||  ||| | |||||  || | |||||||  |||
Db   2326 AGCCTGCCCCTCACCCTGGATCCACCAGGGACAGGG 2291


RESULT   7
US-09-315-372-9/c
; Sequence 9, Application US/09315372
; Patent No. 6057158
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar-Singh, Rajendra
;    APPLICANT:  Hartigan-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States Of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/315,372
```

```
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/735,609
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM-02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3364 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-09-315-372-9


  Query Match             50.0%;  Score 20;  DB 3;  Length 3364;
  Best Local Similarity   72.2%;  Pred. No. 11;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      5 agccaccccataaccctcaatactccagggattggg 40
          ||||  ||| | |||||  || | |||||||  |||
Db   2326 AGCCTGCCCCTCACCCTGGATCCACCAGGGACAGGG 2291


RESULT   8
US-09-244-752-9/c
; Sequence 9, Application US/09244752
; Patent No. 6063622
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar-Singh, Rajendra
;    APPLICANT:  Hartigan-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States Of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,752
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/735,609
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM-02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3364 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-09-244-752-9


  Query Match             50.0%;  Score 20;  DB 3;  Length 3364;
  Best Local Similarity   72.2%;  Pred. No. 11;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      5 agccaccccataaccctcaatactccagggattggg 40
          ||||  ||| | |||||  || | |||||||  |||
Db   2326 AGCCTGCCCCTCACCCTGGATCCACCAGGGACAGGG 2291


RESULT   9
US-09-245-497-9/c
; Sequence 9, Application US/09245497
; Patent No. 6083750
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar-Singh, Rajendra
;    APPLICANT:  Hartigan-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States Of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/245,497
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/735,609
;      APPLICATION NUMBER:  <B) FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM-02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3364 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-09-245-497-9


  Query Match             50.0%;  Score 20;  DB 3;  Length 3364;
  Best Local Similarity   72.2%;  Pred. No. 11;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      5 agccaccccataaccctcaatactccagggattggg 40
          ||||  ||| | |||||  || | |||||||  |||
Db   2326 AGCCTGCCCCTCACCCTGGATCCACCAGGGACAGGG 2291
```

```
RESULT  10
US-09-221-235-10
; Sequence 10, Application US/09221235
; Patent No. 6043040
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,235
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE:
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1864
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (275)..(754)
US-09-221-235-10


  Query Match           48.0%;  Score 19.2;  DB 3;  Length 1864;
  Best Local Similarity  67.5%;  Pred. No. 21;
  Matches   27;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
          ||| |||   |||| | | || ||| |||||   | ||||
Db    955 ctccagcattcccaaagctcttaattctccataaaatggg 994


RESULT  11
US-09-221-928-10
; Sequence 10, Application US/09221928
; Patent No. 6121030
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,928
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE:
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1864
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (275)..(754)
US-09-221-928-10


  Query Match           48.0%;  Score 19.2;  DB 3;  Length 1864;
  Best Local Similarity  67.5%;  Pred. No. 21;
  Matches   27;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
          ||| |||   |||| | | || ||| |||||   | ||||
Db    955 ctccagcattcccaaagctcttaattctccataaaatggg 994


RESULT  12
US-09-221-527-10
; Sequence 10, Application US/09221527
; Patent No. 6146832
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,527
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE:
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1864
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (275)..(754)
US-09-221-527-10


  Query Match           48.0%;  Score 19.2;  DB 3;  Length 1864;
  Best Local Similarity  67.5%;  Pred. No. 21;
  Matches   27;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
          ||| |||   |||| | | || ||| |||||   | ||||
Db    955 ctccagcattcccaaagctcttaattctccataaaatggg 994


RESULT  13
US-09-221-236-10
; Sequence 10, Application US/09221236
; Patent No. 6146841
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,236
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE: 1998-09-29
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1864
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (275)..(754)
US-09-221-236-10


  Query Match           48.0%;  Score 19.2;  DB 3;  Length 1864;
  Best Local Similarity  67.5%;  Pred. No. 21;
  Matches   27;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
          ||| |||   |||| | | || ||| |||||   | ||||
Db    955 ctccagcattcccaaagctcttaattctccataaaatggg 994


RESULT  14
US-09-221-416-10
; Sequence 10, Application US/09221416
; Patent No. 6153417
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,416
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE: 1998-09-29
```

```
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1864
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (275)..(754)
US-09-221-416-10


  Query Match             48.0%;  Score 19.2;  DB 3;  Length 1864;
  Best Local Similarity   67.5%;  Pred. No. 21;
  Matches   27;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 ctcgagccaccccataaccctcaatactccagggattggg 40
          ||| |||   |||| | | || ||| |||||   | ||||
Db    955 ctccagcattcccaaagctcttaattctccataaaatggg 994


RESULT  15
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20


  Query Match             47.0%;  Score 18.8;  DB 2;  Length 32207;
  Best Local Similarity   76.7%;  Pred. No. 54;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      9 accccataaccctcaatactccagggattg 38
          |||  || ||| || | ||||||||||| ||
Db   3464 ACCATATCACCATCCACACTCCAGGGACTG 3493


Search completed: January 18, 2001, 10:59:56
Job time: 30474 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 03:53:45 ; Search time 3453.92 Seconds
(without alignments)
81.154 Million cell updates/sec

Title: PCT-US00-32259-12
Perfect score: 40
Sequence: 1 ctcgagccaccccataaccctcaatactccagggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*
44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estro14:*
121: em_estro15:*
122: em_estro16:*
123: em_estro17:*
124: em_estro18:*
125: em_estro19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est75:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: gb_gss7:*
159: gb_gss8:*
160: gb_gss9:*
161: em_gss5:*
162: em_gss6:*
163: em_gss7:*
164: em_gss8:*
165: em_gss9:*
166: em_gss10:*
167: em_gss11:*
168: gb_gss10:*
169: gb_gss11:*
170: em_gss12:*
171: gb_gss12:*
172: gb_gss13:*
173: gb_gss14:*
174: gb_gss15:*
175: gb_gss16:*
176: gb_gss17:*
177: gb_gss18:*
178: gb_gss19:*
179: em_gss13:*
180: gb_gss20:*
181: gb_gss21:*
182: gb_gss22:*
183: gb_gss23:*
184: gb_gss24:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*
190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	24.6	61.5	296	96	AW845395	AW845395 CM4-CT004
	2	23.8	59.5	250	145	T08487	T08487 EST06378 In
c	3	23.6	59.0	337	146	T92497	T92497 ye24b09.r1
c	4	23	57.5	69	137	BE878784	BE878784 601493062
c	5	23	57.5	118	95	AW804230	AW804230 PM3-UM008
	6	23	57.5	135	133	BE005490	BE005490 CM1-BN011
	7	23	57.5	147	133	BE005492	BE005492 CM1-BN011
c	8	23	57.5	161	143	N73434	N73434 yz31e12.r1
c	9	23	57.5	223	134	BE089733	BE089733 QV0-BT070
	10	23	57.5	227	133	BE005486	BE005486 CM1-BN011
c	11	23	57.5	231	5	AA327421	AA327421 EST30756
c	12	23	57.5	237	139	BF037734	BF037734 601462081
c	13	23	57.5	249	95	AW804233	AW804233 PM3-UM008
c	14	23	57.5	251	110	BE622136	BE622136 601440862
c	15	23	57.5	252	136	BE869246	BE869246 601445061
c	16	23	57.5	252	137	BE894967	BE894967 601436172
c	17	23	57.5	260	2	AA101534	AA101534 zn79c11.s
c	18	23	57.5	263	6	AA383010	AA383010 EST96488
c	19	23	57.5	265	137	BE878754	BE878754 601493017
	20	23	57.5	276	110	BE617770	BE617770 601441968
c	21	23	57.5	289	145	T39447	T39447 ya06a08.r1
c	22	23	57.5	299	104	BE178852	BE178852 PM1-HT060
c	23	23	57.5	304	6	AA384401	AA384401 EST97914
	24	23	57.5	305	12	AA853875	AA853875 NHTBCae09
	25	23	57.5	307	95	AW804241	AW804241 PM3-UM008
c	26	23	57.5	307	145	T39458	T39458 ya06b07.r2
c	27	23	57.5	308	6	AA383479	AA383479 EST96859
c	28	23	57.5	308	95	AW804238	AW804238 PM3-UM008
	29	23	57.5	313	137	BE936437	BE936437 RC1-NT003
c	30	23	57.5	326	6	AA361993	AA361993 EST71440
	31	23	57.5	328	95	AW804271	AW804271 PM3-UM008
c	32	23	57.5	329	135	BE763332	BE763332 RC1-NT003
c	33	23	57.5	332	6	AA373764	AA373764 EST85818
c	34	23	57.5	334	144	R66220	R66220 yi34a07.r1
c	35	23	57.5	335	145	T52898	T52898 ya81f12.r1
c	36	23	57.5	337	38	AV703130	AV703130 AV703130
c	37	23	57.5	342	6	AA363971	AA363971 EST74461
	38	23	57.5	343	137	BE936435	BE936435 RC1-NT003
c	39	23	57.5	344	95	AW804222	AW804222 PM3-UM008
c	40	23	57.5	348	134	BE089684	BE089684 QV0-BT070
c	41	23	57.5	348	134	BE089688	BE089688 QV0-BT070
c	42	23	57.5	349	134	BE089689	BE089689 QV0-BT070
	43	23	57.5	353	133	BE003833	BE003833 QV3-BN009
c	44	23	57.5	354	96	AW904630	AW904630 RC1-NN106
	45	23	57.5	361	133	BE005530	BE005530 CM1-BN011

ALIGNMENTS

RESULT 1
AW845395/c
LOCUS AW845395 296 bp mRNA EST 19-MAY-2000
DEFINITION CM4-CT0045-180200-512-a09 CT0045 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW845395
VERSION AW845395.1 GI:7940912
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 296)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=CM4-CT0045-180
            200-512-a09&t3=2000-02-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 46
            High quality sequence stop: 200.
FEATURES              Location/Qualifiers
     source           1. .296
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="CT0045"
                      /dev_stage="Adult"
                      /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT       70 a     62 c     76 g     87 t      1 others
ORIGIN

  Query Match             61.5%;  Score 24.6;  DB 96;  Length 296;
  Best Local Similarity   76.9%;  Pred. No. 9;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      2 tcgagccacccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||||||| ||||||||||
Db    279 TCCACCCACACTTGAAGCCACAATACTGCAGGGATTGGG 241


RESULT   2
T08487
LOCUS       T08487        250 bp    mRNA            EST       03-AUG-1993
DEFINITION  EST06378 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBBE96
            5' end, mRNA sequence.
ACCESSION   T08487
VERSION     T08487.1  GI:389515
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 250)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nature Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .250
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):85011"
                      /db_xref="taxon:9606"
                      /clone="HIBBE96"
                      /clone_lib="Infant Brain, Bento Soares"
BASE COUNT       56 a     84 c     72 g     31 t      7 others
ORIGIN

  Query Match             59.5%;  Score 23.8;  DB 145;  Length 250;
  Best Local Similarity   77.8%;  Pred. No. 18;
  Matches   28;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      4 gagccacccccataaccctcaatactccagggattgg 39
          ||  | ||||| ||||| |||  || ||||||||||
Db     12 GACACCCCCCAAAACCCCCAACTCTNCAGGGATTGG 47


RESULT   3
T92497/c
LOCUS       T92497        337 bp    mRNA            EST       22-MAR-1995
DEFINITION  ye24b09.r1 Stratagene lung (#937210) Homo sapiens cDNA clone
            IMAGE:118649 5', mRNA sequence.
ACCESSION   T92497
VERSION     T92497.1  GI:724410
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: -97
            Insert Size: 331
            High quality sequence stops: 324
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: M13RP1
            High quality sequence stop: 324.
FEATURES              Location/Qualifiers
     source           1. .337
                      /organism="Homo sapiens"
                      /db_xref="GDB:486938"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:118649"
                      /clone_lib="Stratagene lung (#937210)"
                      /sex="male"
                      /dev_stage="72 years"
                      /lab_host="SOLR cells (kanamycin resistant)"
                      /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
```

```
                    ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                    dT. normal lung.  Average insert size: 1.0 kb; Uni-ZAP XR
                    Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
                    adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       73 a     74 c     94 g     94 t      2 others
ORIGIN


  Query Match             59.0%;  Score 23.6;  DB 146;  Length 337;
  Best Local Similarity   74.4%;  Pred. No. 23;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db    186 TCNACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 148


RESULT   4
BE878784/c
LOCUS       BE878784       69 bp    mRNA            EST       27-SEP-2000
DEFINITION  601493062F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3895430 5',
            mRNA sequence.
ACCESSION   BE878784
VERSION     BE878784.1  GI:10327560
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69)
  AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM673  row: k  column: 15
            High quality sequence stop: 69.
FEATURES             Location/Qualifiers
     source          1. .69
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3895430"
                     /clone_lib="NIH_MGC_69"
                     /tissue_type="large cell carcinoma, undifferentiated"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
BASE COUNT       10 a     16 c     27 g     16 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 137;  Length 69;
  Best Local Similarity   74.4%;  Pred. No. 34;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db     56 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 18


RESULT   5
AW804230/c
LOCUS       AW804230       118 bp    mRNA            EST       16-MAY-2000
DEFINITION  PM3-UM0089-170300-002-d06 UM0089 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW804230
VERSION     AW804230.1  GI:7856100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 118)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=PM3-UM0089-170
            300-002-d06&t3=2000-03-17&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 118.
FEATURES             Location/Qualifiers
     source          1. .118
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="UM0089"
                     /dev_stage="Adult"
                     /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       25 a     31 c     35 g     27 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 95;  Length 118;
  Best Local Similarity   74.4%;  Pred. No. 36;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db     68 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 30


RESULT   6
BE005490
LOCUS       BE005490       135 bp    mRNA            EST       05-JUN-2000
DEFINITION  CM1-BN0117-110400-183-g08 BN0117 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE005490
VERSION     BE005490.1  GI:8265723
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 135)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
```

```
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM1-BN0117-110
            400-183-g08&t3=2000-04-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 18
            High quality sequence stop: 135.
FEATURES             Location/Qualifiers
     source          1. .135
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BN0117"
                     /dev_stage="Adult"
                     /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       29 a     40 c     37 g     29 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 133;  Length 135;
  Best Local Similarity   74.4%;  Pred. No. 36;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db     62 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 100


RESULT   7
BE005492
LOCUS       BE005492      147 bp    mRNA            EST       05-JUN-2000
DEFINITION  CM1-BN0117-110400-183-h04 BN0117 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE005492
VERSION     BE005492.1  GI:8265725
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM1-BN0117-110
            400-183-h04&t3=2000-04-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 147.
FEATURES             Location/Qualifiers
     source          1. .147
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BN0117"
                     /dev_stage="Adult"
                     /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       34 a     43 c     39 g     31 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 133;  Length 147;
  Best Local Similarity   74.4%;  Pred. No. 36;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db     74 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 112


RESULT   8
N73434/c
LOCUS       N73434        161 bp    mRNA            EST       30-JAN-1997
DEFINITION  yz31e12.r1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA
            clone IMAGE:284686 5', mRNA sequence.
ACCESSION   N73434
VERSION     N73434.1  GI:1230719
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1021   Std Error: 0.00
            Seq primer: reverse ET
            High quality sequence stop: 145.
FEATURES             Location/Qualifiers
```

```
    source          1. .161
                    /organism="Homo sapiens"
                    /db_xref="GDB:3903214"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:284686"
                    /clone_lib="Soares_multiple_sclerosis_2NbHMSP"
                    /sex="male"
                    /tissue_type="multiple sclerosis lesions"
                    /dev_stage="Age 46"
                    /lab_host="DH10B (ampicillin resistant)"
                    /note="Vector: pT7T3D (Pharmacia) with a modified
                    polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI
                    ; 1st strand cDNA was primed with a Not I - oligo(dT)
                    primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'],
                    double-stranded cDNA was size selected, ligated to Eco RI
                    adapters (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of a modified pT7T3 vector
                    (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by Bento
                    Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
                    lesions from one patient was kindly provided by Dr. Kevin
                    G. Becker (NINDS/NIH). "
BASE COUNT       36 a     45 c     44 g     36 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 143;  Length 161;
  Best Local Similarity   74.4%;  Pred. No. 37;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db     91 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 53


RESULT   9
BE089733/c
LOCUS       BE089733      223 bp    mRNA            EST       12-JUN-2000
DEFINITION  QV0-BT0704-120500-224-c08 BT0704 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE089733
VERSION     BE089733.1  GI:8480168
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 223)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV0-BT0704-120
            500-224-c08&t3=2000-05-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 223.
FEATURES            Location/Qualifiers
    source          1. .223
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="BT0704"
                    /dev_stage="Adult"
                    /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                    SmaI; A mini-library was made by cloning products derived
                    from  ORESTES PCR (U.S. Letters Patent application No. 196
                    ,716 -  Ludwig Institute for Cancer Research) profiles
                    into  the pUC 18 vector. Reverse transcription of tissue
                    mRNA and cDNA amplification were performed under low
                    stringency conditions."
BASE COUNT       47 a     52 c     70 g     54 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 134;  Length 223;
  Best Local Similarity   74.4%;  Pred. No. 38;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db     86 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 48


RESULT  10
BE005486
LOCUS       BE005486      227 bp    mRNA            EST       05-JUN-2000
DEFINITION  CM1-BN0117-110400-183-e09 BN0117 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE005486
VERSION     BE005486.1  GI:8265719
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 227)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM1-BN0117-110
            400-183-e09&t3=2000-04-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 227.
FEATURES            Location/Qualifiers
    source          1. .227
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="BN0117"
                    /dev_stage="Adult"
                    /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                    Site_2: SmaI; A mini-library was made by cloning products
                    derived from  ORESTES PCR (U.S. Letters Patent application
                    No. 196,716 -  Ludwig Institute for Cancer Research)
```

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 54 a 69 c 54 g 50 t
ORIGIN

Query Match 57.5%; Score 23; DB 133; Length 227;
Best Local Similarity 74.4%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccccataaccctcaatactccagggattggg 40
|| | |||| | || || ||| ||| ||||||||||
Db 154 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 192

RESULT 11
AA327421/c
LOCUS AA327421 231 bp mRNA EST 20-APR-1997
DEFINITION EST30756 Colon I Homo sapiens cDNA 5' end similar to similar to TEGT, mRNA sequence.
ACCESSION AA327421
VERSION AA327421.1 GI:1979666
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 231)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .231
/organism="Homo sapiens"
/db_xref="ATCC (inhost):128102"
/db_xref="taxon:9606"
/clone_lib="Colon I"
/dev_stage="adult"
/note="Organ: colon; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 49 a 50 c 64 g 68 t
ORIGIN

Query Match 57.5%; Score 23; DB 5; Length 231;
Best Local Similarity 74.4%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccccataaccctcaatactccagggattggg 40
|| | |||| | || || ||| ||| ||||||||||
Db 182 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 144

RESULT 12
BF037734/c
LOCUS BF037734 237 bp mRNA EST 10-OCT-2000
DEFINITION 601462081F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3865269 5', mRNA sequence.
ACCESSION BF037734
VERSION BF037734.1 GI:10746117
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 237)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM595 row: b column: 22
High quality sequence stop: 237.
FEATURES Location/Qualifiers
source 1. .237
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3865269"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies. "
BASE COUNT 53 a 48 c 62 g 74 t
ORIGIN

Query Match 57.5%; Score 23; DB 139; Length 237;
Best Local Similarity 74.4%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccccataaccctcaatactccagggattggg 40
|| | |||| | || || ||| ||| ||||||||||
Db 224 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 186

RESULT 13
AW804233/c
LOCUS AW804233 249 bp mRNA EST 16-MAY-2000
DEFINITION PM3-UM0089-170300-002-e01 UM0089 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW804233
VERSION AW804233.1 GI:7856103
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 249)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=PM3-UM0089-170
            300-002-e01&t3=2000-03-17&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 249.
FEATURES             Location/Qualifiers
     source          1. .249
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="UM0089"
                     /dev_stage="Adult"
                     /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       57 a     54 c     73 g     65 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 95;  Length 249;
  Best Local Similarity   74.4%;  Pred. No. 38;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db     84 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 46


RESULT  14
BE622136/c
LOCUS       BE622136      251 bp    mRNA            EST       24-AUG-2000
DEFINITION  601440862F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3915807 5',
            mRNA sequence.
ACCESSION   BE622136
VERSION     BE622136.1  GI:9893076
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 251)
  AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM726  row: l  column: 16
            High quality sequence stop: 251.
FEATURES             Location/Qualifiers
     source          1. .251
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3915807"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT       58 a     51 c     67 g     75 t
ORIGIN


  Query Match             57.5%;  Score 23;  DB 110;  Length 251;
  Best Local Similarity   74.4%;  Pred. No. 38;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 tcgagccaccccataaccctcaatactccagggattggg 40
          || | |||| |   || || ||| ||| ||||||||||
Db    238 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 200


RESULT  15
BE869246/c
LOCUS       BE869246      252 bp    mRNA            EST       27-SEP-2000
DEFINITION  601445061F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3849302 5',
            mRNA sequence.
ACCESSION   BE869246
VERSION     BE869246.1  GI:10318022
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM553  row: i  column: 15
            High quality sequence stop: 252.
FEATURES             Location/Qualifiers
     source          1. .252
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3849302"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
```

BASE COUNT 58 a 50 c 68 g 76 t
ORIGIN

Query Match 57.5%; Score 23; DB 136; Length 252;
Best Local Similarity 74.4%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy       2 tcgagccaccccataaccctcaatactccagggattggg 40
           || | |||| |   || || ||| ||| ||||||||||
Db     239 TCCACCCACACTTGAAGCCACAAAACTGCAGGGATTGGG 201
```

Search completed: January 18, 2001, 03:53:47
Job time: 8901 sec